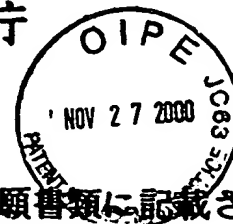


500P0982 UN00

日 本 国 特 許 庁

PATENT OFFICE  
JAPANESE GOVERNMENT



別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日

Date of Application:

2000年 6月 8日

出 願 番 号

Application Number:

特願2000-171556

出 願 人

Applicant (s):

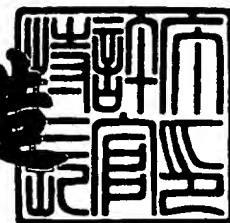
ソニー株式会社

CERTIFIED COPY OF  
PRIORITY DOCUMENT

2000年 8月11日

特許庁長官  
Commissioner,  
Patent Office

及 川 耕 造



出証番号 出証特2000-3063003

特 2000-171556

【書類名】 特許願

【整理番号】 9900951804

【提出日】 平成12年 6月 8日

【あて先】 特許庁長官殿

【国際特許分類】 G11B 20/00

【発明者】

【住所又は居所】 東京都品川区北品川6丁目7番35号 ソニー株式会社  
内

【氏名】 水藤 太郎

【発明者】

【住所又は居所】 東京都品川区北品川6丁目7番35号 ソニー株式会社  
内

【氏名】 太田 正志

【発明者】

【住所又は居所】 東京都品川区北品川6丁目7番35号 ソニー株式会社  
内

【氏名】 三浦 雅美

【特許出願人】

【識別番号】 000002185

【氏名又は名称】 ソニー株式会社

【代表者】 出井 伸之

【代理人】

【識別番号】 100091546

【弁理士】

【氏名又は名称】 佐藤 正美

【電話番号】 03-5386-1775

【先の出願に基づく優先権主張】

【出願番号】 平成11年特許願第236412号

【出願日】 平成11年 8月24日

【手数料の表示】

【予納台帳番号】 048851

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9710846

【ブルーフの要否】 要

【書類名】 明細書

【発明の名称】 音声再生方法および音声再生装置

【特許請求の範囲】

【請求項 1】

記録媒体から、これに記録された音声信号を定常速度より速い速度で再生することによって得られた再生入力音声信号を、処理単位期間ごとに区切り、

出力バッファから定常速度相当量の再生出力音声信号が得られる範囲内で無音部分を削除し、その前後の有音部分を接合して、それぞれの処理単位期間の再生入力音声信号を、前記出力バッファに書き込むとともに、

処理単位期間の再生入力音声信号中に前記出力バッファに蓄えられない量の有音部分が存在する場合には、一部の有音部分を削除し、その前後の有音部分を接合して、その処理単位期間の再生入力音声信号を圧縮して、前記出力バッファに書き込む音声再生方法。

【請求項 2】

請求項 1 の音声再生方法において、

無音部分を削除し、その前後の有音部分を接合するに当たっては、接合点直前の有音部分をフェードアウト処理し、接合点直後の有音部分をフェードイン処理することを特徴とする音声再生方法。

【請求項 3】

請求項 1 の音声再生方法において、

処理単位期間の再生入力音声信号を圧縮するに当たっては、フェード期間において前後の有音部分を重ね合わせるとともに、そのフェード期間における前後の有音部分の差分が最小となるように前後の有音部分の位置関係を調整した上で、クロスフェードによって前後の有音部分を接合することを特徴とする音声再生方法。

【請求項 4】

請求項 1 の音声再生方法において、

処理単位期間を固定の時間とすることを特徴とする音声再生方法。

【請求項 5】

請求項 1 の音声再生方法において、

再生入力音声信号の主成分の音程に応じて処理単位期間の時間を変化させることを特徴とする音声再生方法。

【請求項 6】

請求項 1 の音声再生方法において、

再生入力音声信号の無音部分と有音部分の境目、または有音部分中の相対的にレベルが低い部分を、処理単位期間の区切れ目とすることを特徴とする音声再生方法。

【請求項 7】

請求項 1 の音声再生方法において、

複数チャンネルの音声出力を得る場合に、有音部分と無音部分の識別判定は各チャンネルの再生入力音声信号の平均値によって行い、有音部分の接合は各チャンネルの再生入力音声信号ごとに独立に行うことを特徴とする音声再生方法。

【請求項 8】

請求項 1 の音声再生方法において、

定常再生時、前記出力バッファを映像出力と音声出力を同期させるための時間調整用に用いることを特徴とする音声再生方法。

【請求項 9】

記録媒体から、これに記録された音声信号を定常速度より速い速度で再生することによって得られた再生入力音声信号につき、連続音声部分の終端部の振幅を抑圧し、

その振幅抑圧処理後の再生入力音声信号を、処理単位期間ごとに区切り、

出力バッファから定常速度相当量の再生出力音声信号が得られる範囲内で無音部分を削除し、その前後の有音部分を接合して、それぞれの処理単位期間の再生入力音声信号を、前記出力バッファに書き込むとともに、

処理単位期間の再生入力音声信号中に前記出力バッファに蓄えられない量の有音部分が存在する場合には、一部の有音部分を削除し、その前後の有音部分を接合して、その処理単位期間の再生入力音声信号を圧縮して、前記出力バッファに書き込む音声再生方法。

【請求項 1 0】

請求項 9 の音声再生方法において、

連続音声部分の終端部の振幅を抑圧する処理は、連続音声部分の終端を予測し、その予測した点から振幅の抑圧を開始するとともに、次の連続音声部分の始端が検出されたら、振幅の抑圧を解除することを特徴とする音声再生方法。

【請求項 1 1】

請求項 1 0 の音声再生方法において、

再生入力音声信号の信号レベルが減少傾向にあり、かつ再生入力音声信号の信号レベルが閾値より小さくなったとき、連続音声部分が終端に至る過程に入ったとして、連続音声部分の終端を予測することを特徴とする音声再生方法。

【請求項 1 2】

請求項 1 1 の音声再生方法において、

再生入力音声信号の信号レベルのピーク値を検出し、その検出されたピーク値に応じて前記閾値を設定することを特徴とする音声再生方法。

【請求項 1 3】

請求項 9 の音声再生方法において、

連続音声部分の終端部の振幅を抑圧する処理は、再生入力音声信号を遅延させるとともに、遅延前の再生入力音声信号から連続音声部分の終端を検出し、その検出結果に基づいて遅延後の再生入力音声信号の振幅を抑圧することを特徴とする音声再生方法。

【請求項 1 4】

記録媒体から、これに記録された音声信号を定常速度より速い速度で再生することによって得られた再生入力音声信号に対する速度変換処理の処理単位期間を設定する処理単位期間設定部と、

再生出力音声信号を得るための出力バッファと、

この出力バッファから定常速度相当量の再生出力音声信号が得られる範囲内で無音部分を削除し、その前後の有音部分を接合して、それぞれの処理単位期間の再生入力音声信号を、前記出力バッファに書き込むとともに、処理単位期間の再生入力音声信号中に前記出力バッファに蓄えられない量の有音部分が存在する場

合には、一部の有音部分を削除し、その前後の有音部分を接合して、その処理単位期間の再生入力音声信号を圧縮して、前記出力バッファに書き込む速度変換処理部と、

を備える音声再生装置。

【請求項 1 5】

請求項 1 4 の音声再生装置において、

前記速度変換処理部は、無音部分を削除し、その前後の有音部分を接合するに当たって、接合点直前の有音部分をフェードアウト処理し、接合点直後の有音部分をフェードイン処理することを特徴とする音声再生装置。

【請求項 1 6】

請求項 1 4 の音声再生装置において、

前記速度変換処理部は、処理単位期間の再生入力音声信号を圧縮するに当たって、フェード期間において前後の有音部分を重ね合わせるとともに、そのフェード期間における前後の有音部分の差分が最小となるように前後の有音部分の位置関係を調整した上で、クロスフェードによって前後の有音部分を接合することを特徴とする音声再生装置。

【請求項 1 7】

請求項 1 4 の音声再生装置において、

前記処理単位期間設定部は、処理単位期間を固定の時間とすることを特徴とする音声再生装置。

【請求項 1 8】

請求項 1 4 の音声再生装置において、

当該音声再生装置は、さらに再生入力音声信号の主成分の音程を検出する音程検出部を備え、

前記処理単位期間設定部は、この音程検出部の検出結果に基づいて、再生入力音声信号の主成分の音程に応じて処理単位期間の時間を変化させることを特徴とする音声再生装置。

【請求項 1 9】

請求項 1 4 の音声再生装置において、

前記処理単位期間設定部は、再生入力音声信号の無音部分と有音部分の境目、または有音部分中の相対的にレベルが低い部分を、処理単位期間の区切れ目とすることを特徴とする音声再生装置。

【請求項 2 0】

記録媒体から、これに記録された音声信号を定常速度より速い速度で再生することによって得られた再生入力音声信号につき、連続音声部分の終端部の振幅を抑圧する振幅抑圧処理部と、

その振幅抑圧処理後の再生入力音声信号に対する速度変換処理の処理単位期間を設定する処理単位期間設定部と、

再生出力音声信号を得るための出力バッファと、

この出力バッファから定常速度相当量の再生出力音声信号が得られる範囲内で無音部分を削除し、その前後の有音部分を接合して、それぞれの処理単位期間の再生入力音声信号を、前記出力バッファに書き込むとともに、処理単位期間の再生入力音声信号中に前記出力バッファに蓄えられない量の有音部分が存在する場合には、一部の有音部分を削除し、その前後の有音部分を接合して、その処理単位期間の再生入力音声信号を圧縮して、前記出力バッファに書き込む速度変換処理部と、

を備える音声再生装置。

【発明の詳細な説明】

【0 0 0 1】

【発明の属する技術分野】

この発明は、ビデオテープレコーダなどのように映像信号および音声信号を記録再生する機器や、ミニディスクプレーヤなどのように音声信号を記録再生する機器などにおいて、記録媒体から、これに記録された音声信号を定常速度より速い速度で再生する方法および装置に関する。

【0 0 0 2】

【従来の技術】

家庭用ビデオテープレコーダで、再生時間を短縮するために、テープに記録された映像信号および音声信号を、定常速度より速い 2 倍速などの速度で再生する



場合でも、音声信号については、無音部分を優先的に削除するように元の再生音声信号を周期的に切り貼りすることによって、音声を定常速度で聞き取ることができるようにしたものが考えられている。

【 0 0 0 3 】

【発明が解決しようとする課題】

しかしながら、上述した従来の音声再生方法では、元の再生音声信号を周期的に削除するため、元の再生音声信号中に無音部分が一定の割合で含まれていない場合には、有音部分が削除されて、切り貼り後の再生音声信号は音声が途切れ途切れになり、内容を理解できなくなってしまう。

【 0 0 0 4 】

また、元の再生音声信号中に無音部分がある程度含まれていて、有音部分を削除しなくても再生音声信号の切り貼りができる場合でも、無音期間が短くなり、有音部分が密着することによって、異なる音に聞こえてしまうことがある。例えば「k」のような閉鎖子音の前の無音期間が短くなると、「a - k a」が「a - g a」のように聞こえてしまう。

【 0 0 0 5 】

そこで、この発明は、第 1 に、再生時間を短縮するために定常速度より速い速度で再生された音声信号を通常の音程で出力する場合に、有音部分の削除によって音声の途切れしてしまうことが少なくなり、内容が理解しやすくなるようにしたものである。

【 0 0 0 6 】

この発明は、第 2 に、上記の第 1 に加えて、連続音声部分の密着によって異なる音に聞こえてしまうことが少なくなるようにしたものである。

【 0 0 0 7 】

【課題を解決するための手段】

第 1 の発明の音声再生方法では、

記録媒体から、これに記録された音声信号を定常速度より速い速度で再生することによって得られた再生入力音声信号を、処理単位期間ごとに区切り、

出力バッファから定常速度相当量の再生出力音声信号が得られる範囲内で無音

部分を削除し、その前後の信号部分を接合して、それぞれの処理単位期間の再生入力音声信号を、前記出力バッファに書き込むとともに、

処理単位期間の再生入力音声信号中に前記出力バッファに蓄えられない量の有音部分が存在する場合には、一部の有音部分を削除し、その前後の有音部分を接合して、その処理単位期間の再生入力音声信号を圧縮して、前記出力バッファに書き込む。

【 0 0 0 8 】

第 2 の発明の音声再生方法では、

記録媒体から、これに記録された音声信号を定常速度より速い速度で再生することによって得られた再生入力音声信号につき、連続音声部分の終端部の振幅を抑圧し、

その振幅抑圧処理後の再生入力音声信号を、処理単位期間ごとに区切り、

出力バッファから定常速度相当量の再生出力音声信号が得られる範囲内で無音部分を削除し、その前後の有音部分を接合して、それぞれの処理単位期間の再生入力音声信号を、前記出力バッファに書き込むとともに、

処理単位期間の再生入力音声信号中に前記出力バッファに蓄えられない量の有音部分が存在する場合には、一部の有音部分を削除し、その前後の有音部分を接合して、その処理単位期間の再生入力音声信号を圧縮して、前記出力バッファに書き込む。

【 0 0 0 9 】

上記の方法による第 1 の発明の音声再生方法によれば、出力バッファを効果的に用いることによって、必要最小限のメモリを用いるだけで、有音部分の削除によって音声途切れてしまうことが少なくなり、内容が理解しやすくなる。

【 0 0 1 0 】

第 2 の発明の音声再生方法によれば、さらに、連続音声部分の密着によって異なる音に聞こえてしまうことが少なくなる。

【 0 0 1 1 】

【発明の実施の形態】

〔システム全体の一実施形態…図 1 および図 2〕

図1は、この発明の再生装置の一実施形態を示し、ハードディスク、光磁気ディスク、光ディスクなど、ディスク状の記録媒体から、これに記録された映像信号および音声信号を、定常速度より速い速度で再生することができる場合である。

## 【0012】

再生装置は、記録装置を兼ねてもよく、実際上も記録装置を兼ねることが望ましいが、記録方法および記録装置は、公知のものと特に変わらないので省略し、記録媒体1には、所定のプロトコルおよびフォーマットによるデジタル化された映像信号および音声信号が記録されているものとする。

## 【0013】

以下の例では、MPEG (Moving Picture Experts Group) - 2 Systemsの圧縮符号化方式および多重化方式によって、圧縮された映像データおよび音声データがTS (Transport Stream) として多重化されて記録されているものとし、音声信号のサンプリング周波数は48kHz (60m秒で2880サンプル) とする。

## 【0014】

記録媒体1は、駆動モータ3によって回転駆動される。再生ヘッド2は、記録媒体1から、これに記録されている映像信号および音声信号を読み取るもので、磁気ヘッドや光学ピックアップなどであり、駆動モータ4を含む移動機構によって記録媒体1の径方向に移動させられる。駆動モータ3および4は、サーボ回路5によって駆動制御され、サーボ回路5は、装置全体を制御するシステムコントローラ6によって制御される。

## 【0015】

再生ヘッド2の出力として得られる、記録媒体1から読み取られた映像信号および音声信号、この例ではTSとして多重化された映像データおよび音声データは、デマルチプレクサ7に供給され、デマルチプレクサ7から、映像データおよび音声データが、それぞれ映像PES (Packetized Elementary Stream) および音声PESとして分離されて得られる。

## 【0016】

デマルチプレクサ 7 からの映像データは、MPEG 映像デコーダ 1 1 で、伸長復号化されるとともに、再生倍率に応じてフレーム単位またはフィールド単位で間引かれる。

## 【 0 0 1 7 】

すなわち、2 倍速再生の場合には、記録媒体 1 からは映像データが定常速度の 2 倍の速度で読み取られるが、MPEG 映像デコーダ 1 1 で、連続する 2 フレームにつき 1 フレーム、または連続する 2 フィールドにつき 1 フィールド、映像データが間引かれる。3 倍速再生の場合には、記録媒体 1 からは映像データが定常速度の 3 倍の速度で読み取られるが、MPEG 映像デコーダ 1 1 で、連続する 3 フレームにつき 2 フレーム、または連続する 3 フィールドにつき 2 フィールド、映像データが間引かれる。

## 【 0 0 1 8 】

MPEG 映像デコーダ 1 1 の出力の映像データは、NTSC (National Television System Committee) エンコーダ 1 2 で、NTSC 映像信号にエンコードされ、アナログ信号に変換されて、CRT ディスプレイや液晶ディスプレイなどの表示装置 1 3 に供給される。

## 【 0 0 1 9 】

デマルチプレクサ 7 からの音声データは、MPEG 音声デコーダ 1 4 で、伸長復号化されて、音声処理部 1 5 に供給され、音声処理部 1 5 で、後述のように速度変換されて、再生倍率にかかわらずサンプリング周波数が上記の 4 8 k H z とされ、アナログ信号に変換されて、スピーカなどの音声出力装置 1 6 に供給される。

## 【 0 0 2 0 】

すなわち、2 倍速再生の場合には、記録媒体 1 から音声データが定常速度の 2 倍の速度で読み取られることによって、MPEG 音声デコーダ 1 4 の出力の音声データは、6 0 m 秒で  $2 \times 2880$  サンプルとなるが、音声処理部 1 5 での速度変換によって、音声処理部 1 5 の出力の音声データとしては、6 0 m 秒で 2 8 8 0 サンプルとされる。3 倍速再生の場合には、記録媒体 1 から音声データが定常速度の 3 倍の速度で読み取られることによって、MPEG 音声デコーダ 1 4 の出

力の音声データは、60m秒で3×2880サンプルとなるが、音声処理部15での速度変換によって、音声処理部15の出力の音声データとしては、60m秒で2880サンプルとされる。

## 【0021】

システムコントローラ6には、操作部9が接続される。操作部9は、装置の利用者が再生倍率を指示するなどの各種操作を行うもので、これには、装置の動作状態や操作状況を表示する液晶表示部などの表示部が設けられる。

## 【0022】

図2は、音声処理部15の一例を示す。この例は、MPEG音声デコーダ14の出力の音声データにつき、速度変換の前処理として、連続音声部分（音声のひとまとまり、ひとまとまりの音声部分）の終端部の振幅を抑圧するとともに、速度変換処理の処理単位時間を一定にする場合で、音声処理部15は、振幅抑圧処理部70、処理単位期間設定部21、有音無音判定部22、速度変換処理部23、出力バッファ24、D/Aコンバータ25および音声増幅器26によって構成される。

## 【0023】

〔振幅抑圧処理…図3～図12〕

図2の例の音声処理部15の振幅抑圧処理部70では、MPEG音声デコーダ14の出力の音声データにつき、連続音声部分の終端部の振幅が抑圧される。

## 【0024】

（振幅抑圧処理の第1の例…図3～図6）

振幅抑圧処理部70での振幅抑圧処理の一つの方法として、特開平8-179792号に示された方法を用いることができる。図3～図6に、この特開平8-179792号に示された方法による振幅抑圧処理を示す。

## 【0025】

図3は、振幅抑圧処理部70の構成を示し、入力端子71には、入力音声信号として上述したMPEG音声デコーダ14の出力の音声データが供給される。この入力音声信号は、子音成分分離フィルタ72およびフォルマント成分分離フィルタ73に供給され、子音成分分離フィルタ72からは、入力音声信号中の子音

成分が抽出される。また、フォルマント成分分離フィルタ 7 3 の通過帯域が、例えば 1 5 0 ~ 1 0 0 0 H z に設定されることによって、フォルマント成分分離フィルタ 7 3 からは、入力音声信号中のピッチ成分およびフォルマント成分が抽出される。

## 【 0 0 2 6 】

フォルマント成分分離フィルタ 7 3 の出力は、レベル検出部 7 4 に供給され、レベル検出部 7 4 において、例えば、フォルマント成分分離フィルタ 7 3 の出力が全波整流され、その整流出力が 6 0 H z 以下を通過帯域とするローパスフィルタに供給され、ローパスフィルタの出力のレベルが検出されることによって、入力音声信号の音声レベルが検出され、レベル検出値 E が得られる。このようにピッチ成分およびフォルマント成分のみから音声レベルが検出されることによって、レベル検出値 E としてノイズの影響が少ないものが得られる。

## 【 0 0 2 7 】

このレベル検出部 7 4 からのレベル検出値 E は、制御部 7 5 で後述のように処理され、制御部 7 5 から振幅抑圧の制御信号としての制御係数 W が得られる。

## 【 0 0 2 8 】

そして、フォルマント成分分離フィルタ 7 3 の出力、および制御部 7 5 からの制御係数 W が、振幅抑圧部 7 6 に供給され、振幅抑圧部 7 6 において、後述のように制御係数 W によってフォルマント成分分離フィルタ 7 3 の出力が減衰される。

## 【 0 0 2 9 】

さらに、周波数特性補正フィルタ 7 7 で、子音成分分離フィルタ 7 2 の出力と振幅抑圧部 7 6 の出力が混合されるとともに、混合後の信号に対してイコライジング処理などの所要の周波数特性補正処理が施され、処理後の出力音声信号が出力端子 7 9 に得られる。周波数特性補正フィルタ 7 7 は、倍速再生の再生倍率に応じてフィルタ係数や処理帯域が設定される。ただし、周波数特性補正フィルタ 7 7 によって周波数特性を補正しないで、子音成分分離フィルタ 7 2 の出力と振幅抑圧部 7 6 の出力を混合するだけでもよい。

## 【 0 0 3 0 】

制御部 7 5 は、レベル検出部 7 4 からのレベル検出値  $E$  から、連続音声部分の終端が近づいていることを予測するとともに、次の連続音声部分の始端を検出する。そして、その予測および検出に基づいて、振幅抑圧部 7 6 に対する制御係数  $W$  を設定して、連続音声部分の終端が近づいたら、フォルマント成分分離フィルタ 7 3 の出力を減衰させ、次の連続音声部分の始端が到来したら、その減衰を解除することによって、隣接する連続音声部分の間に振幅の小さい期間または無音期間を形成し、または隣接する連続音声部分の間の無音期間を広げる。

## 【 0 0 3 1 】

この処理を、図 4 を用いて示す。同図 (A) は、入力端子 7 1 に得られる入力音声信号であり、アナログ波形で示す。同図 (B) は、この場合にレベル検出部 7 4 から得られるレベル検出値  $E$  であり、アナログ波形で示す。

## 【 0 0 3 2 】

制御部 7 5 は、まず、レベル検出値  $E$  のピーク値  $P K$  を検出して閾値  $T H$  を設定する。例えば、ピーク値  $P K$  に係数を乗じて、ピーク値  $P K$  の一定割合の値を閾値  $T H$  として設定する。次に、そのときのサンプルタイミングでのレベル検出値  $E(i)$  を直前のサンプルタイミングでのレベル検出値  $E(i-1)$  と比較して、音声レベルが増加傾向にあるか減少傾向にあるかを判断するとともに、そのときのサンプルタイミングでのレベル検出値  $E(i)$  を上記の閾値  $T H$  と比較する。

## 【 0 0 3 3 】

そして、図 4 の時点  $t_0$  で示すように、音声レベルが減少傾向にあり、かつ、そのときのレベル検出値  $E(i)$  が閾値  $T H$  より小さくなったときには、連続音声部分の終端が近づいていると判断して、同図 (C) に示すように、振幅抑圧部 7 6 に対する制御係数  $W$  (振幅抑圧部 7 6 の利得レベル) を最大値 1 から最小値 0 に向けて徐々に低下させ、フォルマント成分分離フィルタ 7 3 の出力の減衰量を徐々に増加させる。

## 【 0 0 3 4 】

また、振幅抑圧後、図 4 の時点  $t_1$  で示すように、音声レベルが増加傾向に転じたときには、次の連続音声部分の始端と判断して、同図 (C) に示すように、

振幅抑圧部 7 6 に対する制御係数  $W$  (振幅抑圧部 7 6 の利得レベル) を最小値 0 から最大値 1 に向けて徐々に増加させ、フォルマント成分分離フィルタ 7 3 の出力の減衰量を徐々に低下させる。

## 【 0 0 3 5 】

このような振幅抑圧処理によって、出力端子 7 9 に得られる出力音声信号は、図 4 (D) にアナログ波形で示すように、隣接する連続音声部分の間に振幅の小さい期間または無音期間が形成され、または隣接する連続音声部分の間の無音期間が広げられたものとなる。

## 【 0 0 3 6 】

制御部 7 5 は、より具体的には、図 5 に示すような処理ルーチンによって、振幅抑圧部 7 6 に対する制御係数  $W$  を設定する。

## 【 0 0 3 7 】

図 5 の処理ルーチンは、レベル検出部 7 4 から制御部 7 5 にレベル検出値  $E(i)$  が取り込まれるごとに処理を開始するもので、まずステップ F 1 0 1 で、レベル検出値  $E(i)$  を閾値  $TH$  と比較する。閾値  $TH$  は、それ以前においてステップ F 1 1 0 または F 1 1 7 で後述のように設定した値である。

## 【 0 0 3 8 】

そして、 $E(i) \geq TH$  であれば、ステップ F 1 0 1 からステップ F 1 1 1 に進んで、ダウンフラグ  $DW$  を 1 にセットする。ダウンフラグ  $DW$  は、それ以前は音声レベルが減少傾向にあった場合、または減少傾向にあったと見なす場合に、1 とするフラグである。

## 【 0 0 3 9 】

次に、ステップ F 1 1 2 で、フラグ  $BY$  を確認する。フラグ  $BY$  は、直前のレベル検出値  $E(i-1)$  が閾値  $TH$  より小さかった場合に、ステップ F 1 0 2 で 0 とされるフラグである。したがって、レベル検出値  $E(i)$  として閾値  $TH$  以上の値が最初に入力されたときには、フラグ  $BY$  は 0 であって、ステップ F 1 1 2 からステップ F 1 1 8 に進む。

## 【 0 0 4 0 】

ステップ F 1 1 8 では、レベル検出値  $E(i)$  をピーク値  $PK$  の下限値  $Lim$



Lと比較する。ピーク値PKについては、あらかじめ上限値LimHおよび下限値LimLが設定され、制御部75は、後述のように、ピーク値PKが上限値LimHより大きいときには、上限値LimHをピーク値PKとして設定し、ピーク値PKが下限値LimLより小さいときには、下限値LimLをピーク値PKとして設定して、ピーク値PKを上限値LimHと下限値LimLとの間に設定するものである。

## 【0041】

そして、 $E(i) < LimL$ であれば、ステップF118からステップF119に進んで、下限値LimLをピーク値PKとして設定した上で、ステップF121に進み、 $E(i) \geq LimL$ であれば、ステップF118からステップF120に進んで、閾値THをピーク値PKとして設定した上で、ステップF121に進む。

## 【0042】

ステップF121では、上記のフラグBYを1にセットし、次にレベル検出値E(i)が入力されたときに備えて、音声レベルが閾値TH以上になったことを示しておく。

## 【0043】

次に、ステップF122で、そのときのレベル検出値E(i)を直前のレベル検出値E(i-1)と比較し、 $E(i) > E(i-1)$ であれば、音声レベルが増加傾向にあると判断して、ステップF122からステップF123に進んで、制御係数Wを $W + d2$ の値に更新した上で、次回の処理に移行し、 $E(i) \leq E(i-1)$ であれば、音声レベルが増加傾向にないと判断して、ステップF122から直接、次回の処理に移行する。

## 【0044】

ステップF121からステップF122に進んだ場合は、レベル検出値E(i)として閾値TH以上の値が最初に入力されたときで、音声レベルが増加傾向にあるので、ステップF123で制御係数Wが更新されることになる。

## 【0045】

d2は、制御係数Wを増加させ、減衰量を低下させる場合のステップ幅である

。ただし、図 4 (C) に示したように、制御係数  $W$  は  $0 \leq W \leq 1$  とされる。したがって、 $W = 1$  であったときには、ステップ F 1 2 3 では制御係数  $W$  を更新しない。

## 【 0 0 4 6 】

次回の処理においても、 $E(i) \geq TH$  であれば、ステップ F 1 0 1 からステップ F 1 1 1 および F 1 1 2 に進むが、このときは、レベル検出値  $E(i)$  として閾値  $TH$  以上の値が最初に入力されたときではなく、フラグ  $BY$  が 1 にセットされているので、ステップ F 1 1 2 からステップ F 1 1 3 に進む。

## 【 0 0 4 7 】

ステップ F 1 1 3 では、レベル検出値  $E(i)$  をピーク値  $PK$  と比較し、 $PK \geq E(i)$  であれば、そのままステップ F 1 2 2 に進むが、 $PK < E(i)$  であれば、ステップ F 1 1 4 以下に進んで、ピーク値  $PK$  を更新した上で、閾値  $TH$  を更新する。

## 【 0 0 4 8 】

すなわち、まずステップ F 1 1 4 で、レベル検出値  $E(i)$  をピーク値  $PK$  の上限値  $LimH$  と比較し、 $E(i) > LimH$  であれば、ステップ F 1 1 4 からステップ F 1 1 5 に進んで、上限値  $LimH$  をピーク値  $PK$  として設定した上で、ステップ F 1 1 7 に進み、 $E(i) \leq LimH$  であれば、ピーク値  $PK$  より大きいレベル検出値  $E(i)$  が新たなピーク値  $PK$  として検出されたことになるので、ステップ F 1 1 4 からステップ F 1 1 6 に進んで、レベル検出値  $E(i)$  をピーク値  $PK$  として設定した上で、ステップ F 1 1 7 に進む。

## 【 0 0 4 9 】

ステップ F 1 1 7 では、閾値  $TH$  を、上記のように更新したピーク値  $PK$  の一定割合の値  $PK \times RT$  に更新する。係数  $RT$  で表される割合は、数 % ~ 数 10 % に選定される。ステップ F 1 1 7 で閾値  $TH$  を更新したら、ステップ F 1 2 2 に進む。

## 【 0 0 5 0 】

以上は、レベル検出値  $E(i)$  が閾値  $TH$  以上の値である場合である。逆に、レベル検出値  $E(i)$  が閾値  $TH$  より小さい場合には、ステップ F 1 0 1 からス

テップ F 1 0 2 に進んで、上記のフラグ B Y を 0 にする。

【 0 0 5 1 】

次に、ステップ F 1 0 3 で、そのときのレベル検出値  $E(i)$  を直前のレベル検出値  $E(i-1)$  と比較し、 $E(i) < E(i-1)$  であれば、音声レベルが減少傾向にあると判断して、ステップ F 1 0 3 からステップ F 1 0 5 に進んで、ダウンフラグ D W を確認する。

【 0 0 5 2 】

そして、ダウンフラグ D W が 1 であれば、それ以前は音声レベルが減少傾向にあった場合、または減少傾向にあったと見なす場合であるので、このとき、レベル検出値  $E(i)$  が閾値 T H より小さく、かつ直前のレベル検出値  $E(i-1)$  より小さいことから、連続音声部分の終端が近づいていると判断して、ステップ F 1 0 5 からステップ F 1 0 6 に進んで、制御係数 W を  $W-d1$  の値に更新した上で、次の処理に移行する。

【 0 0 5 3 】

d 1 は、制御係数 W を減少させ、減衰量を増加させる場合のステップ幅である。ただし、図 4 (C) に示したように、制御係数 W は  $0 \leq W \leq 1$  とされる。したがって、 $W=0$  であったときには、ステップ F 1 0 6 では制御係数 W を更新しない。

【 0 0 5 4 】

以後、レベル検出値  $E(i)$  が連続音声部分の終端に向けて低下し続けている間は、ステップ F 1 0 6 で制御係数 W の更新が繰り返されて、図 4 の時点  $t_0$  以降に示したように、振幅抑圧部 7 6 の利得レベルが徐々に低下する。そして、制御係数 W (利得レベル) が 0 に達すると、次に時点  $t_1$  で示したように音声レベルが増加傾向に転じるまでは、制御係数 W (利得レベル) が 0 の状態が維持される。

【 0 0 5 5 】

一方、ステップ F 1 0 5 でダウンフラグ D W が 0 であると判断した場合は、それ以前は音声レベルが減少傾向になく、レベル検出値  $E(i)$  として閾値 T H より小さい値が最初に入力されたときである。このとき、音声レベルが減少傾向に

あっても、連続音声部分の終端が近づいていない場合も存在すると考えられる。そこで、ステップF105でダウンフラグDWが0であると判断したときには、ステップF107以下に進んで、ピーク値PKを更新した上で、閾値THを更新する。

## 【0056】

すなわち、まずステップF107で、レベル検出値 $E(i)$ をピーク値PKの下限值 $LimL$ と比較し、 $E(i) \geq LimL$ であれば、ステップF107からステップF108に進んで、レベル検出値 $E(i)$ をピーク値PKとして設定した上で、ステップF110に進み、 $E(i) < LimL$ であれば、ステップF107からステップF109に進んで、下限値 $LimL$ をピーク値PKとして設定した上で、ステップF110に進む。

## 【0057】

ステップF110では、閾値THを、上記のように更新したピーク値PKの一定割合の値 $PK \times RT$ に更新する。また、このとき、連続音声部分の終端が近づいている場合も存在するので、ステップF106と同様に、制御係数Wを $W-d1$ の値に更新する。さらに、ステップF110では、ダウンフラグDWを1にセットし、音声レベルが減少傾向になったことを示しておく。

## 【0058】

このようにレベル検出値 $E(i)$ として閾値THより小さい値が最初に入力されたときには、ステップF110で閾値THが更新されることによって、連続音声部分の終端が近づいているか否かが、より正確に判定される。すなわち、実際には連続音声部分の終端が近づいていないときには、次に入力されるレベル検出値 $E(i)$ は更新された閾値TH以上になる確率が高く、次の処理ではステップF111以降の処理が実行される可能性が高いので、連続音声部分の終端が近づいていると誤判定される確率は低い。

## 【0059】

逆に、実際に連続音声部分の終端が近づいているときには、次に入力されるレベル検出値 $E(i)$ が更新された閾値THより小さくなり、かつ直前のレベル検出値 $E(i-1)$ より小さくなるので、次の処理ではステップF105からス

テップ F106 に進んで、連続音声部分の終端部の振幅を抑圧する処理が実行される。

#### 【0060】

振幅抑圧後、図4の時点  $t_1$  で示したように次の連続音声部分の始端が到来すると、制御部75は、ステップ F103 で、そのときのレベル検出値  $E(i)$  が直前のレベル検出値  $E(i-1)$  以上であると判断することによって、ステップ F103 からステップ F104 に進んで、制御係数  $W$  を  $W+d_2$  の値に更新するとともに、ダウンフラグ  $DW$  を 0 にする。

#### 【0061】

以後、レベル検出値  $E(i)$  が増加する間は、レベル検出値  $E(i)$  が閾値  $TH$  より小さいときにはステップ F104 で、レベル検出値  $E(i)$  が閾値  $TH$  以上の値のときにはステップ F123 で、制御係数  $W$  の更新が繰り返されて、図4の時点  $t_1$  以降に示したように、振幅抑圧部76の利得レベルが徐々に増加する。

#### 【0062】

以上のように、図5の処理ルーチンによって、振幅抑圧処理部70の出力音声信号は、隣接する連続音声部分の間に振幅の小さい期間または無音期間が形成され、または隣接する連続音声部分の間の無音期間が広げられたものとなる。

#### 【0063】

図6は、話者が日本語の「親譲りの無鉄砲で子供のときから損ばかりしている」という発音をしたときの音声信号波形を示し、同図(A)が、振幅抑圧処理を行う前の波形であり、同図(B)が、上述した振幅抑圧処理を行った後の波形である。同図(B)の矢印で示すように、上述した振幅抑圧処理を行った場合には、それぞれの話音(連続音声部分)の間が振幅の小さい期間または無音期間によって明確に区切られ、それぞれの話音を明瞭に認識することができるようになる。

#### 【0064】

(振幅抑圧処理の第2の例…図7～図12)

図2の例の音声処理部15の振幅抑圧処理部70での振幅抑圧処理の他の一つ

の方法として、特開平 7-3 6 4 8 7 号に示された方法を用いることができる。  
図 7～図 1 2 に、この特開平 7-3 6 4 8 7 号に示された方法による振幅抑圧処理を示す。

## 【 0 0 6 5 】

図 7 は、振幅抑圧処理部 7 0 の構成を示し、入力端子 7 1 には、入力音声信号として上述した M P E G 音声デコーダ 1 4 の出力の音声データが供給される。この入力音声信号は、遅延回路 8 1 で例えば 3 0 m 秒遅延されて、振幅抑圧部 8 2 に供給される。また、入力音声信号が、バンドパスフィルタ 8 3 に供給されて、バンドパスフィルタ 8 3 から、入力音声信号中のピッチ成分および第 1 フォルマント成分のみが抽出される。

## 【 0 0 6 6 】

人の会話の音声は、母音の場合、主としてピッチ成分と第 1 フォルマント成分、第 2 フォルマント成分などのフォルマント成分とに解析することができる。すなわち、例えば「あ」と発音したときの音声パワースペクトルは、図 9 に示すように、周波数が低い方から順に、ピッチ成分 P、第 1 フォルマント成分 F 1、第 2 フォルマント成分 F 2、第 3 フォルマント成分 F 3 …と、エネルギーが集中する箇所が存在する。

## 【 0 0 6 7 】

したがって、バンドパスフィルタ 8 3 の上側のカットオフ周波数を第 1 フォルマント成分 F 1 と第 2 フォルマント成分 F 2 との間の周波数に選定することによって、ピッチ成分および第 1 フォルマント成分のみを抽出することができる。ただし、発音する音や発音する者の違いによって各成分の周波数が変化するので、システムが扱う音声信号によってバンドパスフィルタ 8 3 の通過帯域を若干変える必要がある。

## 【 0 0 6 8 】

このバンドパスフィルタ 8 3 からのピッチ成分および第 1 フォルマント成分は、レベル検出部 8 4 に供給され、レベル検出部 8 4 において、例えば、バンドパスフィルタ 8 3 の出力が全波整流され、その整流出力が 6 0 H z 以下を通過帯域とするローパスフィルタに供給され、ローパスフィルタの出力のレベルが検出さ

れることによって、入力音声信号の音声レベルが検出され、レベル検出値Eが得られる。このようにピッチ成分およびフォルマント成分のみから音声レベルが検出されることによって、レベル検出値Eとしてノイズの影響が少ないものが得られる。

## 【 0 0 6 9 】

このレベル検出部 8 4 からのレベル検出値Eは、制御部 8 5 で後述のように処理され、振幅抑圧部 8 2 において、後述のように制御部 8 5 の制御によって遅延回路 8 1 の出力の音声信号が減衰され、出力端子 7 9 に振幅抑圧処理後の出力音声信号が得られる。

## 【 0 0 7 0 】

図 1 0 は、この例の制御部 8 5 が実行する処理ルーチンの一例を示す。この処理ルーチンは、レベル検出部 8 4 から制御部 8 5 にレベル検出値E ( i ) が取り込まれるごとに処理を開始するもので、まずステップ S 1 0 1 で、レベル検出値E ( i ) を音声の途切れを判断するための閾値A t h 1 と比較する。閾値A t h 1 は、それ以前においてステップ S 1 0 7 で後述のように設定した値である。

## 【 0 0 7 1 】

そして、 $E(i) \geq Ath1$ であれば、ステップ S 1 0 1 からステップ S 1 0 2 に進んで、イネーブル値が 0 であるか否かを判断する。このイネーブル値は、以前の音声レベルの状態を示す値で、以前に音声レベルが増加している場合には 1 とされるものである。

## 【 0 0 7 2 】

そのため、音声レベルが立ち上がり始めたときには、ステップ S 1 0 2 ではイネーブル値が 0 であると判断して、ステップ S 1 0 2 からステップ S 1 0 3 に進んで、レベル検出値E ( i ) を音声の存在を判断するための閾値A t h 2 と比較する。閾値A t h 2 は、あらかじめ設定される。

## 【 0 0 7 3 】

そして、 $E(i) \leq Ath2$ であれば、そのまま次回の処理に移行するが、 $E(i) > Ath2$ であれば、音声が存在すると判断して、ステップ S 1 0 4 に進んで、上記のイネーブル値を 1 に設定するとともに、振幅抑圧終了時からのサン

ブル数を示す変数  $S_o$  を 0 に設定した上で、ステップ S105 に進む。

【0074】

ステップ S105 では、レベル検出値  $E(i)$  をピーク値  $PK$  の上限値  $LimH$  と比較し、 $E(i) > LimH$  であれば、ステップ S105 からステップ S106 に進んで、上限値  $LimH$  をピーク値  $PK$  として設定した上で、ステップ S107 に進み、 $E(i) \leq LimH$  であれば、ステップ S105 からステップ S108 に進んで、レベル検出値  $E(i)$  をピーク値  $PK$  として設定した上で、ステップ S107 に進む。

【0075】

ステップ S107 では、音声の途切れを判断するための閾値  $Ath1$  を、上記のように設定したピーク値  $PK$  の一定割合の値  $PK \times Rt$  に更新する。ステップ S107 で閾値  $Ath1$  を更新したら、次の処理に移行する。

【0076】

一方、ステップ S102 でイネーブル値が 1 であると判断したときには、ステップ S109 に進んで、レベル検出値  $E(i)$  をピーク値  $PK$  と比較し、 $E(i) \leq PK$  であれば、そのまま次の処理に移行するが、 $E(i) > PK$  であれば、ステップ S105 以下に進んで、上記のようにピーク値  $PK$  および閾値  $Ath1$  を更新した上で、次の処理に移行する。

【0077】

以上の処理を繰り返している間にレベル検出値  $E(i)$  が低下して、ステップ S101 でレベル検出値  $E(i)$  が閾値  $Ath1$  より小さいと判断したときには、ステップ S101 からステップ S110 に進んで、上記のイネーブル値が 1 であるか否かを判断する。

【0078】

そして、イネーブル値が 1 であれば、連続音声部分の終端であると判断して、ステップ S110 からステップ S111 に進んで、イネーブル値を 0 に変更し、振幅抑圧終了時からのサンプル数を示す変数  $S_o$  を 0 に設定するとともに、利得漸減期間（減衰量漸増期間）および無音期間を形成するように振幅抑圧部 82 の利得レベルを制御する。



## 【 0 0 7 9 】

すなわち、図 1 1 に示すようなサンプル数  $S_a$  の期間  $T_a$  では、遅延回路 8 1 の出力の音声信号に対する重み係数（振幅抑圧部 8 2 の利得レベル）を、最大値 1 から最小値 0 に向けて、係数  $W_a$  で示すように音声信号の 1 サンプルごとに徐々に低下させ、その直後の、図 1 1 に示すようなサンプル数  $M$  の期間  $T_m$  では、遅延回路 8 1 の出力の音声信号に対する重み係数（振幅抑圧部 8 2 の利得レベル）を最小値 0 にする。利得漸減期間  $T_a$  は、例えば 1 0 m 秒とし、無音期間  $T_m$  は、例えば 2 0 m 秒とする。

## 【 0 0 8 0 】

このようにステップ  $S_{111}$  で振幅抑圧部 8 2 の利得レベルを制御したら、次の処理に移行する。次の処理でも、ステップ  $S_{101}$  でレベル検出値  $E(i)$  が閾値  $A_{th1}$  より小さいと判断するので、ステップ  $S_{101}$  からステップ  $S_{110}$  に進んで、イネーブル値が 1 であるか否かを判断するが、次の処理では、イネーブル値が 0 にされているので、ステップ  $S_{110}$  からステップ  $S_{112}$  に進む。

## 【 0 0 8 1 】

そして、ステップ  $S_{112}$  では、振幅抑圧終了時からのサンプル数を示す変数  $S_o$  が、利得漸増期間（減衰量漸減期間）として定められる期間のサンプル数  $S_b$  より小さいか否かを判断し、 $S_o < S_b$  であれば、次の連続音声部分の始端であると判断して、ステップ  $S_{112}$  からステップ  $S_{113}$  に進んで、利得漸増期間を形成するように振幅抑圧部 8 2 の利得レベルを制御する。

## 【 0 0 8 2 】

すなわち、図 1 1 に示すような、無音期間  $T_m$  の直後のサンプル数  $S_b$  の期間  $T_b$  では、遅延回路 8 1 の出力の音声信号に対する重み係数（振幅抑圧部 8 2 の利得レベル）を、最小値 0 から最大値 1 に向けて、係数  $W_b$  で示すように音声信号の 1 サンプルごとに徐々に増加させる。利得漸増期間  $T_b$  は、例えば 5 m 秒とする。

## 【 0 0 8 3 】

このようにステップ  $S_{113}$  で振幅抑圧部 8 2 の利得レベルを制御したら、次

回の処理に移行する。ステップ S 1 1 2 で変数 S o がサンプル数 S b 以上であると判断したときには、そのまま次の処理に移行する。

#### 【 0 0 8 4 】

以上のように、この例の振幅抑圧処理では、制御部 8 5 の制御によって振幅抑圧部 8 2 で音声信号の振幅が抑圧されるが、振幅抑圧部 8 2 に供給される音声信号は、入力端子 7 1 に得られる音声信号に対して 3 0 m 秒遅れるので、制御部 8 5 で連続音声部分の終端が検出されたとき、3 0 m 秒前に遡って連続音声部分の終端の 3 0 m 秒前の信号位置から、利得漸減期間 T a および無音期間 T m で示したように音声信号の振幅を抑圧することができる。

#### 【 0 0 8 5 】

図 1 2 は、話者が日本語の「書いて下さい」という発音をしたときの音声信号波形を示し、同図 (A) が、振幅抑圧処理を行う前の波形であり、同図 (B) が、上述した振幅抑圧処理を行った後の波形である。同図 (B) に示すように、上述した振幅抑圧処理を行った場合には、それぞれ「かい」「て」「く」「だ」「さい」という話音（連続音声部分）の終端の直前に無音期間 M 1, M 2, M 3, M 4, M 5 が形成され、それぞれの話音を明瞭に認識することができるようになる。

#### 【 0 0 8 6 】

無音期間 T m は、2 0 m 秒に限らず、数 m 秒～数 1 0 m 秒の範囲で適切に設定すればよい。また、完全な無音状態にしないで、例えば振幅を数 1 0 % 抑圧してもよい。

#### 【 0 0 8 7 】

図 8 に示すように、図 7 の遅延回路 8 1 および振幅抑圧部 8 2 の代わりに、F I F O ( f i r s t - i n f i r s t - o u t ) メモリ 8 6 を用いてもよい。この場合には、F I F O メモリ 8 6 に書き込んだ音声データを一定時間遅れて F I F O メモリ 8 6 から読み出す際、制御部 8 5 によってデータ値を補正することによって、出力端子 7 9 に振幅抑圧処理後の出力音声信号を得る。

#### 【 0 0 8 8 】

〔速度変換処理…図 1 3 ～図 1 7 〕

図2の例の音声処理部15では、MPEG音声デコーダ14の出力の音声データにつき、上述したように振幅抑圧処理部70で連続音声部分の終端部の振幅を抑圧した後、速度変換（人の会話の場合には話速変換）を実行する。

## 【0089】

MPEG音声デコーダ14の出力の音声データ、したがって振幅抑圧処理部70の出力の音声データは、上述したように、2倍速再生の場合には60m秒で $2 \times 2880$ サンプルとなり、3倍速再生の場合には60m秒で $3 \times 2880$ サンプルとなる。以下では、2倍速再生の場合を例として速度変換処理を示すが、3倍速再生などの場合でも同様である。

## 【0090】

図2の例の音声処理部15の処理単位期間設定部21は、システムコントローラ6からのタイミング信号およびクロックによって、図13に処理単位期間T1～T6として示すように、一音素の時間より短い60m秒相当（2880サンプル分、2倍速再生の場合には絶対時間としては30m秒）の時間間隔で、速度変換処理のそれぞれの処理単位期間の先頭ないし末尾のタイミングを決定する。ただし、処理単位期間設定部21は、システムコントローラ6の一部、または速度変換処理部23の一部として、構成することもできる。

## 【0091】

有音無音判定部22は、入力データSi、すなわち振幅抑圧処理部70の出力の音声データの、有音部分と無音部分を識別判定するもので、この例では、図14に示すように、入力データSiの、それぞれ60m秒相当（2880サンプル）の処理単位期間を、3m秒相当（144サンプル分、2倍速再生の場合には絶対時間としては1.5m秒）ずつの20個の判定単位期間に分割し、それぞれの判定単位期間につき、144サンプルのデータの絶対値の平均値を算出して、その平均値が閾値を超えるか否かを判断し、平均値が閾値を超えるときには、その判定単位期間を有音部分と判定し、平均値が閾値以下であるときには、その判定単位期間を無音部分と判定する。

## 【0092】

出力バッファ24は、FIFO構造のリングバッファとして機能するものとす

る。図 1 5 に、出力バッファ 2 4 へのデータ書き込み、出力バッファ 2 4 からのデータ読み出し、および出力バッファ 2 4 の入出力ポインタの動きを示す。

## 【 0 0 9 3 】

出力バッファ 2 4 にデータが全く書き込まれていないときには、図 1 5 (A) に示すように、出力バッファ 2 4 には先頭アドレスから順にデータが書き込まれ、それに応じて入力ポインタも後方に移動する。このとき、出力ポインタは先頭アドレスを指示する。

## 【 0 0 9 4 】

図 1 5 (B) は、この状態から、さらにデータが書き込まれるとともに、先頭アドレスから順にデータが読み出された状態を示し、出力ポインタも後方に移動する。

## 【 0 0 9 5 】

さらにデータが書き込まれて、入力ポインタが末尾アドレスに達すると、図 1 5 (C) に示すように、入力ポインタは先頭アドレスに戻る。このとき、新たなデータが出力済みデータに対してのみ上書きされるように、入力ポインタの指示位置が制御される。出力ポインタも、末尾アドレスに達したときには、先頭アドレスに戻る。

## 【 0 0 9 6 】

D/A コンバータ 2 5 は、出力バッファ 2 4 から出力された音声データをアナログ音声信号に変換するものであり、音声増幅器 2 6 は、そのアナログ音声信号を増幅するものである。音声増幅器 2 6 からのアナログ音声信号は、図 1 の音声出力装置 1 6 に供給される。

## 【 0 0 9 7 】

速度変換処理部 2 3 は、図 1 6 に示す速度変換処理ルーチンによって、振幅抑圧処理部 7 0 の出力の音声データ、すなわち入力データ  $S_i$  を速度変換するものである。

## 【 0 0 9 8 】

すなわち、この速度変換処理ルーチン 3 0 では、まずステップ 3 1 において、有音無音判定部 2 2 の判定結果から、入力データ  $S_i$  のそれぞれの処理単位期間

内に無音部分が倍率相当分以上存在するか否かを判断する。倍率相当分とは、処理単位時間から、処理単位時間に再生倍率の逆数を乗じた時間を引いた時間で、上記のように処理単位時間が60m秒相当（2880サンプル）で、かつ2倍速再生の場合には、30m秒分（1440サンプル）である。すなわち、ステップ31では、それぞれの処理単位期間内に無音部分が30m秒分（1440サンプル）以上存在するか否かを判断する。

## 【0099】

そして、処理単位期間内に無音部分が30m秒分（1440サンプル）以上存在するときには、ステップ31からステップ32に進んで、出力バッファ24に30m秒分（1440サンプル）以上のデータが蓄積されているか否かを判断し、出力バッファ24に30m秒分（1440サンプル）以上のデータが蓄積されているときには、ステップ32からステップ33に進んで、その処理単位期間内の無音部分を全て削除し、無音部分の前後の有音部分を接合して、その処理単位期間内の有音部分のデータのみを全て、出力バッファ24に書き込む。

## 【0100】

また、出力バッファ24に30m秒分（1440サンプル）以上のデータが蓄積されていないときには、ステップ32からステップ34に進んで、その処理単位期間内の30m秒分（1440サンプル）以下の有音部分の全てに、その処理単位期間の先頭に近い部分から、その処理単位期間内の無音部分を一部付加して、その処理単位期間内の30m秒分（1440サンプル）のデータを、出力バッファ24に書き込む。

## 【0101】

ただし、処理単位期間内に有音部分と無音部分が、ちょうど30m秒分（1440サンプル）ずつ存在するときには、ステップ33と同様に、その処理単位期間内の無音部分を全て削除し、無音部分の前後の有音部分を接合して、その処理単位期間内の有音部分のデータのみを全て、出力バッファ24に書き込む。

## 【0102】

一方、処理単位期間内に無音部分が30m秒分（1440サンプル）未満しか存在しないときには、すなわち有音部分が30m秒分（1440サンプル）を超

えて存在するときには、ステップ 3 1 からステップ 3 5 に進んで、出力バッファ 2 4 に、その処理単位期間内の 3 0 m 秒分（1 4 4 0 サンプル）を超える有音部分のデータを全て蓄えるだけの空きがあるか否かを判断する。

【 0 1 0 3 】

そして、出力バッファ 2 4 に、その処理単位期間内の 3 0 m 秒分（1 4 4 0 サンプル）を超える有音部分のデータを全て蓄えるだけの空きがあるときには、ステップ 3 5 からステップ 3 3 に進んで、その処理単位期間内の無音部分を全て削除し、無音部分の前後の有音部分を接合して、その処理単位期間内の有音部分のデータのみを全て、出力バッファ 2 4 に書き込む。

【 0 1 0 4 】

ただし、その処理単位期間内の 6 0 m 秒分（2 8 8 0 サンプル）のデータが全て有音部分であるときには、その 6 0 m 秒分（2 8 8 0 サンプル）の有音部分のデータをそのまま、出力バッファ 2 4 に書き込む。

【 0 1 0 5 】

また、出力バッファ 2 4 に、その処理単位期間内の 3 0 m 秒分（1 4 4 0 サンプル）を超える有音部分のデータを全て蓄えるだけの空きがないときには、ステップ 3 5 からステップ 3 6 に進んで、有音部分が 3 0 m 秒分（1 4 4 0 サンプル）となるようにデータを圧縮し、その圧縮後のデータを、出力バッファ 2 4 に書き込む。後述のように、この場合は速聴きとなる。

【 0 1 0 6 】

この圧縮処理について示すと、例えば、図 1 7 の上段に示すように、その処理単位期間内の 6 0 m 秒分（2 8 8 0 サンプル）のデータが全て有音部分であるときには、同図の下段に示すように、中間部分の 1 0 4 0 サンプルのデータを削除し、期間の先頭から 9 2 0 サンプルの前方部データ S f 中の後寄りの 4 0 0 サンプルのデータと、期間の末尾から 9 2 0 サンプルの後方部データ S b 中の前寄りの 4 0 0 サンプルのデータとを、フェード期間として、前方部データ S f と後方部データ S b をクロスフェードによって接合する。このようにクロスフェードによって接合するのは、接合された部分で圧縮後のデータにノイズが発生するのを防止するためである。

## 【 0 1 0 7 】

処理単位期間内に無音部分が存在するときには、その無音部分を含めた 1 0 4 0 サンプルのデータを削除して、それぞれ 9 2 0 サンプルの有音部分を同様のクロスフェードによって接合する。無音部分が 1 4 4 0 サンプル未満ではあるが、1 0 4 0 サンプルを超えるときには、4 0 0 サンプルのフェード期間が得られ、接合後のデータとして 1 4 4 0 サンプルのデータが得られるように、有音部分の一部として無音部分の一部を含ませる。

## 【 0 1 0 8 】

図 1 3 の処理単位期間 T 1 は、入力データ S i 中に破線で囲んだ有音部分と破線で囲まない無音部分が、ちょうど 3 0 m 秒分（1 4 4 0 サンプル）ずつ存在するとともに、再生開始直後であるため、出力バッファ 2 4 にはデータが全く蓄積されていない場合である。

## 【 0 1 0 9 】

したがって、処理単位期間 T 1 については、ステップ 3 1, 3 2 および 3 4 によって、図 1 3 の出力データ S o 中の部分 E 1 に示すように、ちょうど 3 0 m 秒分（1 4 4 0 サンプル）の無音部分が全て削除され、無音部分の前後の有音部分が接合されて、ちょうど 3 0 m 秒分（1 4 4 0 サンプル）の有音部分のデータのみが全て、出力バッファ 2 4 に書き込まれる。

## 【 0 1 1 0 】

ただし、図 1 3 は入力データ S i と出力データ S o のタイミング関係を示すものではなく、タイミング的には出力データ S o は入力データ S i に対して同図の位置より遅れる。

## 【 0 1 1 1 】

出力バッファ 2 4 に書き込まれた処理単位期間 T 1 内の 3 0 m 秒分（1 4 4 0 サンプル）の有音部分のデータは、3 0 m 秒の時間に渡って、出力バッファ 2 4 から読み出され、D/A コンバータ 2 5 によってアナログ音声信号に変換される。

## 【 0 1 1 2 】

図 1 3 の処理単位期間 T 2 は、6 0 m 秒分（2 8 8 0 サンプル）のデータが全

て有音部分であるとともに、出力バッファ 24 に、この 60 m 秒分（2880 サンプル）の有音部分のデータを全て蓄えるだけの空きがある場合である。

## 【0113】

したがって、処理単位期間 T2 については、ステップ 31, 35 および 33 によって、図 13 の出力データ S<sub>o</sub> 中の部分 E2 に示すように、その 60 m 秒分（2880 サンプル）の有音部分のデータがそのまま、出力バッファ 24 に書き込まれる。

## 【0114】

この出力バッファ 24 に書き込まれた処理単位期間 T2 の 60 m 秒分（2880 サンプル）の有音部分のデータ中の、前半の 30 m 秒分（1440 サンプル）のデータが、期間 T1 内の有音部分のデータが読み出される 30 m 秒の時間に続く 30 m 秒の時間に渡って、出力バッファ 24 から読み出され、D/A コンバータ 25 によってアナログ音声信号に変換される。

## 【0115】

図 13 の処理単位期間 T3 は、入力データ S<sub>i</sub> 中に破線で囲んだ有音部分が 30 m 秒分（1440 サンプル）未満しか存在せず、破線で囲まない無音部分が 30 m 秒分（1440 サンプル）を超えて存在するとともに、出力バッファ 24 に 30 m 秒分（1440 サンプル）以上のデータが蓄積されている場合である。

## 【0116】

したがって、処理単位期間 T3 については、ステップ 31, 32 および 33 によって、図 13 の出力データ S<sub>o</sub> 中の部分 E3 に示すように、その 30 m 秒分（1440 サンプル）を超える無音部分が全て削除され、無音部分の前後の有音部分が接合されて、30 m 秒分（1440 サンプル）未満の有音部分のデータのみが全て、出力バッファ 24 に書き込まれる。

## 【0117】

そして、処理単位期間 T2 の 60 m 秒分（2880 サンプル）の有音部分のデータ中の、出力バッファ 24 に蓄積されていた後半の 30 m 秒分（1440 サンプル）のデータが、前半のデータが読み出される 30 m 秒の時間に続く 30 m 秒の時間に渡って、出力バッファ 24 から読み出され、D/A コンバータ 25 によ



ってアナログ音声信号に変換される。

【0118】

図13の処理単位期間T4は、60m秒分（2880サンプル）のデータが全て有音部分であるとともに、出力バッファ24に、この60m秒分（2880サンプル）の有音部分のデータを全て蓄えるだけの空きがある場合である。

【0119】

したがって、処理単位期間T4については、ステップ31、35および33によって、図13の出力データS<sub>o</sub>中の部分E4に示すように、その60m秒分（2880サンプル）の有音部分のデータがそのまま、出力バッファ24に書き込まれる。

【0120】

そして、出力バッファ24に蓄積されていた処理単位期間T3の30m秒分（1440サンプル）未満の有音部分のデータと、出力バッファ24に書き込まれた処理単位期間T4の60m秒分（2880サンプル）の有音部分のデータ中の前半の一部のデータとを合わせた、30m秒分（1440サンプル）のデータが、30m秒の時間に渡って、出力バッファ24から読み出され、D/Aコンバータ25によってアナログ音声信号に変換される。

【0121】

図13の処理単位期間T5は、60m秒分（2880サンプル）のデータが全て有音部分であるとともに、出力バッファ24に、この60m秒分（2880サンプル）の有音部分のデータを全て蓄えるだけの空きがない場合である。

【0122】

したがって、処理単位期間T5については、ステップ31、35および36によって、図13の出力データS<sub>o</sub>中の部分E5に示し、かつ図17に示したように、60m秒分（2880サンプル）の有音部分のデータが、30m秒分（1440サンプル）に圧縮されて、出力バッファ24に書き込まれる。

【0123】

そして、処理単位期間T4の60m秒分（2880サンプル）の有音部分のデータ中の前半の残りの部分のデータと、後半の一部のデータとを合わせた、30

m秒分（1440サンプル）のデータが、30m秒の時間に渡って、出力バッファ24から読み出され、D/Aコンバータ25によってアナログ音声信号に変換される。

#### 【0124】

さらに、処理単位期間T4の60m秒分（2880サンプル）の有音部分のデータ中の後半の残りの部分のデータと、処理単位期間T5の30m秒分（1440サンプル）に圧縮された有音部分のデータ中の一部のデータとを合わせた、30m秒分（1440サンプル）のデータが、30m秒の時間に渡って、出力バッファ24から読み出され、D/Aコンバータ25によってアナログ音声信号に変換される。

#### 【0125】

以上のように、図13の例では、処理単位期間T1～T4については、入力データSi中の無音部分のみが全て削除され、有音部分のデータのみが全て出力バッファ24に書き込まれるとともに、出力バッファ24からは有音部分のデータのみが全て、60m秒で2880サンプルの速度で読み出されるので、出力倍率は1倍となり、有音部分のみが全て定常速度で出力される。

#### 【0126】

処理単位期間T5については、60m秒分（2880サンプル）の有音部分のデータが、30m秒分（1440サンプル）に圧縮されて、出力バッファ24に書き込まれ、30m秒の時間で出力バッファ24から読み出されるので、例外的に有音部分が削減されて、速聴きとなる。

#### 【0127】

なお、図16のステップ34で、有音部分の全てに無音部分を一部付加して、30m秒分（1440サンプル）のデータを出力バッファ24に書き込む場合については、図13には示してしないが、例えば、処理単位期間T2が図13とは異なり、有音部分が30m秒分（1440サンプル）未満しか存在せず、無音部分が30m秒分（1440サンプル）を超えて存在する場合には、図13の処理単位期間T2では、出力バッファ24に30m秒分（1440サンプル）以上のデータが蓄積されていないので、その30m秒分（1440サンプル）未満の有

音部分の全てに、期間 T 2 の先頭に近い部分から、30 m 秒分（1440 サンプル）を超える無音部分の一部が付加されて、期間 T 2 内の 30 m 秒分（1440 サンプル）のデータが、出力バッファ 24 に書き込まれる。

【0128】

以上のように、上述した例によれば、出力バッファ 24 を効果的に用いることによって、必要最小限のメモリを用いるだけで、有音部分の削除によって音声途切れてしまうことが少なくなり、内容が理解しやすくなる。

【0129】

〔速度変換処理の際の音質向上…図 18～図 21〕

（データ圧縮の際の位置調整…図 18、図 19）

図 16 のステップ 36 で、図 13 の部分 E 5 および図 17 の下段に示したようにクロスフェードによって前方部データ S f と後方部データ S b を接合する場合、図 18（A）に示すように、フェード期間において前方部データ S f と後方部データ S b の位相が一致しているときには、接合後のデータは接合部分でうなりを生じない。

【0130】

しかし、図 18（B）に示すように、フェード期間において前方部データ S f と後方部データ S b の位相が大きくずれているときには、接合後のデータは接合部分でうなりを生じ、音質が劣化する。

【0131】

そこで、クロスフェードによって前方部データ S f と後方部データ S b を接合するに当たっては、以下のように前方部データ S f と後方部データ S b の位置関係を調整する。

【0132】

すなわち、図 19 に示すように、後方部データ S b を、前方部データ S f に対して、所定サンプル数の時間内で、1 サンプルまたは数サンプルの時間ずつ、タイミング的に移動させて、それぞれの移動位置において、上述した 400 サンプル分のフェード期間内の前方部データ S f のデータ値と後方部データ S b のデータ値との差分の絶対値の、接合後のデータとして示す波形の塗り潰した部分の面

積で表されるフェード期間全体に渡る積分値（和）を算出する。

【 0 1 3 3 】

そして、この積分値が最小となる移動位置を、前方部データ  $S_f$  と後方部データ  $S_b$  が最適な位相関係になる位置として検出し、その移動位置において、上述したようにクロスフェードによって前方部データ  $S_f$  と後方部データ  $S_b$  を接合する。

【 0 1 3 4 】

前方部データ  $S_f$  および後方部データ  $S_b$  のフェード期間におけるデータを、それぞれ  $F[n]$  および  $B[n]$  とし、フェード期間のサンプル数を  $j$  とすると、接合後のデータの接合部分（フェード期間）におけるデータ  $X[n]$  は、

$$X[n] = ((j - n) * F[n] + n * B[n]) / j \cdots (1)$$

で表される。ただし、 $n = 0, 1, 2 \cdots (j - 1)$  である。

【 0 1 3 5 】

この例によれば、接合後のデータは接合部分でうなりや違和感の少ないものとなる。

【 0 1 3 6 】

（無音部分削除の際のノイズ発生防止…図 2 0、図 2 1）

図 1 6 のステップ 3 4 で、図 1 3 の部分 E 1 に示したように、または図 1 6 のステップ 3 3 で、図 1 3 の部分 E 3 に示したように、入力データ  $S_i$  中の無音部分を削除して、前後の有音部分を接合する場合、無音部分として図 2 0 の上段に示すデータ部分  $S_p$  を削除したときには、部分  $S_p$  は無音部分でレベルが低いものの、図 2 0 の下段左側に示すように、接合後の出力データ  $S_o$  は接合点の前後で位相が大きく変化するため、接合点でポツ音と呼ばれるノイズが発生する。また、無音部分として図 2 0 の上段に示すデータ部分  $S_q$  を削除したときには、図 2 0 の下段右側に示すように、同様に接合後の出力データ  $S_o$  には接合点でポツ音と呼ばれるノイズが発生する。

【 0 1 3 7 】

そこで、削除した無音部分の前後の有音部分を接合するに当たっては、接合点の前後のそれぞれ所定サンプル数  $k$  の有音部分につき、フェードアウト処理およ

びフェードイン処理を行う。

【0138】

具体的に、図20下段の左側および右側に示すように、フェードアウト処理前の接合部前半のデータをC1[i]、フェードイン処理前の接合部後半のデータをD1[i]とし、図21の左側および右側に示すように、フェードアウト処理後の接合部前半のデータをC2[i]、フェードイン処理後の接合部後半のデータをD2[i]とすると、

$$C2[i] = (k - i) * C1[i] / k \quad \dots (2)$$

$$D2[i] = i * D1[i] / k \quad \dots (3)$$

とする。ただし、 $i = 0, 1, 2 \dots (k - 1)$ である。

【0139】

このようにフェードアウト処理およびフェードイン処理をして、前後の有音部分を接合することによって、図21に示すように、接合後の出力データS<sub>o</sub>は接合点でボツ音と呼ばれるノイズが発生しなくなる。

【0140】

〔速度変換処理の処理単位時間を変化させる場合…図22～図26〕

(音程に応じて処理単位時間を変化させる場合…図22、図23)

図22は、図1の音声処理部15の他の例を示し、音声信号の主成分の音程に応じて速度変換処理の処理単位時間を変化させる場合である。音程に対して処理単位時間を直線的に変化させることも可能であるが、以下の例は、音程を3段階に分けて処理単位時間を3段階に変化させる場合である。

【0141】

この例では、音程検出部27において、振幅抑圧処理部70の出力の音声信号の主成分（レベルが最大の周波数成分）の音程が検出される。音程検出部27での音程検出方法には、FFT（高速フーリエ変換）などの公知の方法を用いることができる。

【0142】

そして、この例では、処理単位期間設定部21は、音程検出部27の検出結果に基づいて、図23に示す音程対応処理単位期間設定処理ルーチンによって、処

理単位時間を決定し、処理単位期間の先頭ないし末尾のタイミングを決定する。

【 0 1 4 3 】

すなわち、この音程対応処理単位期間設定処理ルーチン40では、まずステップ41において、音程検出部27の検出結果の音程が300Hz以上であるか否かを判断し、300Hz以上であるときには、ステップ41からステップ42に進んで、処理単位時間を最短の40m秒相当（1920サンプル分で、2倍速再生の場合には絶対時間としては20m秒）として、当該の処理単位期間の先頭（直前の処理単位期間の末尾）から40m秒相当後を当該の処理単位期間の末尾（直後の処理単位期間の先頭）とする。

【 0 1 4 4 】

検出結果の音程が300Hz未満であるときには、ステップ41からステップ43に進んで、検出結果の音程が100Hz以下であるか否かを判断し、100Hz以下であるときには、ステップ43からステップ44に進んで、処理単位時間を最長の60m秒相当（2880サンプル分で、2倍速再生の場合には絶対時間としては30m秒）として、当該の処理単位期間の先頭から60m秒相当後を当該の処理単位期間の末尾とする。

【 0 1 4 5 】

検出結果の音程が100Hz以下でもないときには、すなわち100Hzを超え、かつ300Hz未満であるときには、ステップ43からステップ45に進んで、処理単位時間を中間の50m秒相当（2400サンプル分で、2倍速再生の場合には絶対時間としては25m秒）として、当該の処理単位期間の先頭から50m秒相当後を当該の処理単位期間の末尾とする。

【 0 1 4 6 】

図22の速度変換処理部23が行う速度変換処理は、図13の処理単位期間T1、T2、T3…が、60m秒相当という固定の時間ではなく、40m秒相当、50m秒相当または60m秒相当というように変化し、図16の速度変換処理ルーチン30における倍率相当分が、処理単位時間の変化に応じて、2倍速再生の場合には20m秒分（960サンプル）、25m秒分（1200サンプル）または30m秒分（1440サンプル）というように変化する点を除いて、上述した

例と同じである。

【 0 1 4 7 】

この例によれば、音声信号の主成分の音程が高いときには速度変換処理の処理単位時間が短くなり、主成分の音程が低いときには速度変換処理の処理単位時間が長くなるので、音声信号の周波数による歪みや揺れの少ない出力音声信号を得ることができる。

【 0 1 4 8 】

(低レベルの点を処理単位期間の区切れ目とする場合…図 2 4 ～図 2 6)

図 2 4 は、図 1 の音声処理部 1 5 のさらに他の例を示し、無音部分と有音部分の境目、または有音部分中の相対的にレベルが低い部分を、速度変換処理の処理単位期間の区切れ目とする場合である。

【 0 1 4 9 】

この例では、処理単位期間設定部 2 1 は、振幅抑圧処理部 7 0 の出力の音声信号、および有音無音判定部 2 2 の判定結果を取り込んで、図 2 5 に示す音声レベル対応処理単位期間設定処理ルーチンによって、処理単位期間の先頭ないし末尾のタイミングを決定する。

【 0 1 5 0 】

すなわち、この音声レベル対応処理単位期間設定処理ルーチン 5 0 では、まずステップ 5 1 において、有音無音判定部 2 2 の判定結果から、当該の処理単位期間の先頭から 3 0 m 秒相当以上、1 0 0 m 秒相当以下の時間範囲内に、無音部分と有音部分の切り替わり時点（無音部分から有音部分に切り替わる時点、または有音部分から無音部分に切り替わる時点）が存在するか否かを判断する。

【 0 1 5 1 】

そして、その時間範囲内に無音部分と有音部分の切り替わり時点が存在するときには、ステップ 5 1 からステップ 5 2 に進んで、その無音部分と有音部分の切り替わり時点を当該の処理単位期間の末尾とする。その時間範囲内に無音部分と有音部分の切り替わり時点が 2 点以上存在するときには、最初の時点を当該の処理単位期間の末尾とする。

【 0 1 5 2 】

一方、その時間範囲内に無音部分と有音部分の切り替わり時点が存在しないときには、ステップ 5 1 からステップ 5 3 に進んで、その時間範囲内で音声平均レベルが最小となる時点を当該の処理単位期間の末尾とする。音声平均レベルは、図 1 4 に示したような判定単位期間内の各サンプルのデータ値の絶対値の平均値であり、その平均値が最小となる判定単位期間の先頭または末尾を当該の処理単位期間の末尾とする。

## 【 0 1 5 3 】

したがって、この例では、処理単位時間は、最短で 3 0 m 秒相当、最長で 1 0 0 m 秒相当である。

## 【 0 1 5 4 】

図 2 6 に「固定長の場合」として示すケースは、図 2 の例のような構成で、処理単位期間  $T_1$ 、 $T_2$ 、 $T_3 \dots$  が 6 0 m 秒相当という固定の時間にされる場合であり、図 2 6 に「可変長の場合」として示すケースは、この図 2 4 の例のような構成で、処理単位期間  $T_{11}$ 、 $T_{12}$ 、 $T_{13} \dots$  の先頭ないし末尾のタイミングが決定され、時間を変えられる場合である。ただし、図 2 6 に示す時間は、絶対時間ではなく、上述したように 2 8 8 0 サンプル分を 6 0 m 秒相当としたときの時間である。

## 【 0 1 5 5 】

すなわち、この例では、時間軸上で 0 m 秒として示す最初の処理単位期間  $T_{11}$  の先頭の時点から 3 0 m 秒相当以上、1 0 0 m 秒相当以下の時間範囲内に、入力データ  $S_i$  が有音部分から無音部分に切り替わる時点  $t_a$  が存在するので、その時点  $t_a$  が、処理単位期間  $T_{11}$  の末尾、すなわち次の処理単位期間  $T_{12}$  の先頭となる。

## 【 0 1 5 6 】

さらに、処理単位期間  $T_{12}$  の先頭の時点  $t_a$  から 3 0 m 秒相当以上、1 0 0 m 秒相当以下の時間範囲内に、入力データ  $S_i$  が無音部分から有音部分に切り替わる時点  $t_b$  が存在するので、その時点  $t_b$  が、処理単位期間  $T_{12}$  の末尾、すなわち次の処理単位期間  $T_{13}$  の先頭となる。

## 【 0 1 5 7 】



さらに、処理単位期間 T 1 3 の先頭の時点 t b から 3 0 m 秒相当以上、1 0 0 m 秒相当以下の時間範囲内には、無音部分と有音部分の切り替わり時点が存在しないので、その時間範囲内で入力データ S i の音声平均レベルが最小となる時点 t c が、処理単位期間 T 1 3 の末尾、すなわち次の処理単位期間 T 1 4 の先頭となる。

## 【 0 1 5 8 】

この例によれば、無音部分と有音部分の境目、または有音部分中の相対的にレベルが低い部分が、速度変換処理の処理単位期間の区切れ目となるので、音声信号の切り貼りによる音質の劣化を大幅に低減することができる。

## 【 0 1 5 9 】

〔多チャンネル再生の場合…図 2 7、図 2 8〕

上述した例は、図 1 の記録媒体 1 から 1 チャンネルの音声信号を再生する場合として示した。しかし、記録媒体 1 には多チャンネルの音声信号を記録することができ、その多チャンネルの音声信号からユーザの希望するチャンネルの音声信号を選択して、または多チャンネルの音声信号を同時に、再生することが可能である。以下に、その場合の例を示す。

## 【 0 1 6 0 】

（第 1 の例…図 2 7）

図 2 7 に、ステレオ音声信号、またはバイリンガルなどの 2 元音声（主音声および副音声）中的一方または両方の音声信号を、選択的に再生する場合の例を示す。

## 【 0 1 6 1 】

この例では、システムコントローラ 6 からユーザによって選択された音声モードを示す音声モード識別信号が得られて、有音無音判定部 2 2 および速度変換処理部 2 3 に供給され、ユーザによって選択された音声モードに応じて、有音無音判定部 2 2 での有音無音判定および速度変換処理部 2 3 での速度変換処理が、以下のように切り換え制御される。

## 【 0 1 6 2 】

すなわち、ステレオ音声信号を再生する場合には、M P E G 音声デコーダ 1 4

からは左右のチャンネルの音声データが交互に得られ、有音無音判定部 2 2 は、図 1 4 に示したような判定単位期間のそれぞれにつき、左チャンネルのデータと右チャンネルのデータの平均値を算出して、その平均値が閾値を超えるか否かを判断し、平均値が閾値を超えるときには、その判定単位期間を有音部分と判定し、平均値が閾値以下であるときには、その判定単位期間を無音部分と判定する。

**【 0 1 6 3 】**

速度変換処理部 2 3 は、M P E G 音声デコーダ 1 4 からの左右のチャンネルの音声データを、処理単位期間設定部 2 1 によって設定された処理単位期間ごとに、一旦、左チャンネルのデータと右チャンネルのデータに分離して、内部バッファ 2 9 の左チャンネル用および右チャンネル用として設定した別個のエリアに書き込む。

**【 0 1 6 4 】**

さらに、速度変換処理部 2 3 は、その別個のエリアに書き込んだ左右のチャンネルの音声データを、上記の有音無音判定部 2 2 の判定結果に基づいて、それぞれ図 1 6 に示した速度変換処理ルーチン 3 0 によって独立に速度変換処理し、処理後の左右のチャンネルの音声データを、左チャンネルのデータと右チャンネルのデータが交互に配列されるように出力バッファ 2 4 に書き込む。

**【 0 1 6 5 】**

出力バッファ 2 4 からは、速度変換処理後の左右のチャンネルの音声データを、交互に読み出して、D / A コンバータ 2 5 によってアナログ音声信号に変換し、D / A コンバータ 2 5 からの左右のチャンネルのアナログ音声信号を、音声増幅器 2 6 a および 2 6 b に振り分けて出力する。

**【 0 1 6 6 】**

一方、主音声と副音声のいずれか一方の音声信号を再生する場合には、有音無音判定部 2 2 は、図 1 4 に示したような判定単位期間のそれぞれにつき、選択された方の音声の、すなわち主音声が選択されたときには主音声の、副音声が選択されたときには副音声の、音声データの絶対値の平均値を算出して、その平均値が閾値を超えるか否かによって、有音部分と無音部分を識別判定する。

**【 0 1 6 7 】**

速度変換処理部 2 3 は、選択された方の音声についてのみ、すなわち主音声を選択されたときには主音声についてのみ、副音声を選択されたときには副音声についてのみ、MPEG 音声デコーダ 1 4 の出力の音声データを、上記の有音無音判定部 2 2 の判定結果に基づいて、図 1 6 に示した速度変換処理ルーチン 3 0 によって速度変換処理し、処理後の音声データを出力バッファ 2 4 に書き込む。

## 【 0 1 6 8 】

そして、出力バッファ 2 4 から速度変換処理後の音声データを読み出し、D/A コンバータ 2 5 によってアナログ音声信号に変換して、音声増幅器 2 6 a および 2 6 b に同時に出力する。

## 【 0 1 6 9 】

主音声と副音声の音声信号を同時に再生する場合には、ステレオ音声信号を再生する場合と同様に、有音無音判定部 2 2 は、主音声と副音声の音声データの平均値を算出して、その平均値が閾値を超えるか否かによって、有音部分と無音部分を識別判定し、速度変換処理部 2 3 は、主音声と副音声の音声データを独立に速度変換処理することによって、音声増幅器 2 6 a および 2 6 b から、それぞれ主音声および副音声のアナログ音声信号を出力する。

## 【 0 1 7 0 】

ただし、主音声と副音声の音声信号を同時に再生する場合、有音無音判定部 2 2 で、主音声と副音声の音声データにつき独立に、有音部分と無音部分の識別判定を行ってもよい。

## 【 0 1 7 1 】

図 2 7 の例では、振幅抑圧処理部を省略したが、この例でも、振幅抑圧処理部を設けて、MPEG 音声デコーダ 1 4 の出力の音声データにつき、連続音声部分の終端部の振幅を抑圧することができる。この場合、ステレオ音声信号を再生する場合、または主音声と副音声の音声信号を同時に再生する場合には、速度変換処理と同様に、左右のチャンネルの音声データ、または主音声と副音声の音声データにつき、独立に振幅抑圧処理する。

## 【 0 1 7 2 】

以上は、2 チャンネルの場合であるが、3 チャンネル以上の場合も、同様であ

る。

### 【0173】

(第2の例…図28)

さらに、一般に2チャンネル以上の多チャンネルの場合、図28の例のように構成することもできる。同図の例は、図1の再生装置において、デマルチプレクサ7から多チャンネルの音声データを分離し、それぞれMPEG音声デコーダ14a, 14b, 14c…で伸長復号化して、音声処理部15a, 15b, 15c…に供給するもので、音声処理部15a, 15b, 15c…は、それぞれ図2、図22または図24の例の音声処理部15と同様に構成する。

### 【0174】

この例では、例えば、チャンネル1については、その音声信号を再生するとともに、音声処理部15aの速度変換処理部23aで速度変換処理し、チャンネル2については、その音声信号を再生するものの、音声処理部15bの速度変換処理部23bで速度変換処理しない、というように、チャンネル1, 2, 3…の各チャンネルの音声信号を、選択的または同時に再生し、選択的または同時に速度変換処理することができる。

### 【0175】

〔映像信号および音声信号を定常速度で再生する場合…図29〕

図1の再生装置で、記録媒体1から映像信号および音声信号を定常速度で再生する場合、MPEG音声デコーダ14からは音声データが1倍速で出力され、音声処理部15の図2または図27などに示した処理単位期間設定部21、有音無音判定部22および速度変換処理部23は、それぞれの機能を停止して、音声処理部15が振幅抑圧処理部70を備える場合には、振幅抑圧処理部70の出力の音声データがそのまま、音声処理部15が振幅抑圧処理部70を備えない場合には、MPEG音声デコーダ14の出力の音声データがそのまま、出力バッファ24に書き込まれ、出力バッファ24から読み出される。

### 【0176】

しかし、MPEG映像デコーダ11では、映像データの伸長復号化に少なくとも数フレームの時間を要するので、NTSCエンコーダ12からの映像出力は、

音声処理部 1 5 からの音声出力に対して時間遅れを生じ、映像出力と音声出力の同期が得られなくなる。

【 0 1 7 7 】

そこで、以下の例では、定常再生時には、出力バッファ 2 4 を利用して、この映像出力の音声出力に対する遅れを吸収し、映像出力と音声出力を同期させる。

【 0 1 7 8 】

図 2 9 は、その例を示し、速度変換処理部 2 3 は、システムコントローラ 6 によって指示された再生モードに応じて、出力バッファ 2 4 に制御信号を送出して、出力バッファ 2 4 からデータを読み出すタイミングを制御する。

【 0 1 7 9 】

具体的には、図 1 5 に示したように F I F O 構造のリングバッファとして機能する出力バッファ 2 4 の出力ポインタを制御する。すなわち、出力バッファ 2 4 の出力ポインタと入力ポインタの間に蓄積されるデータ量が、映像出力の音声出力に対する遅れ時間に相当するように、出力ポインタを制御する。

【 0 1 8 0 】

例えば、この例によらないときの映像出力の音声出力に対する遅れ時間を  $D_v$  (m秒)、音声信号のサンプリング周波数を  $F_s$  (k H z)、音声データをデータ幅が  $B_h$  (ビット) のステレオ (2 チャンネル) 音声信号とすると、出力ポインタと入力ポインタの間のデータ蓄積量  $A_d$  (バイト) を、

$$A_d = (D_v \times F_s \times B_h \times 2) / 8 \quad \cdots (4)$$

とする。

【 0 1 8 1 】

以上のように、この例によれば、映像信号と音声信号を定常速度で再生する場合に、出力バッファ 2 4 を利用して映像出力と音声出力を同期させることができる。

【 0 1 8 2 】

〔他の実施形態〕

図 1 の実施形態は、記録媒体 1 がディスク状の記録媒体である場合であるが、この発明は、記録媒体 (記憶媒体) が磁気テープや半導体メモリなどである場合

にも、同様に適用することができる。半導体メモリの場合には、再生ヘッドではなく読み出し回路によってメモリから、これに書き込まれた音声信号を読み出すことは言うまでもない。この場合には、「記録」とは書き込みであり、「再生」とは読み出しである。

【0183】

また、この発明は、映像信号および音声信号を定常速度で再生する場合についての図29のような例を除いて、記録媒体から音声信号のみを再生する場合または装置にも、同様に適用することができる。

【0184】

【発明の効果】

上述したように、第1の発明によれば、再生時間を短縮するために定常速度より速い速度で再生された音声信号を通常の音程で出力する場合に、出力バッファを効果的に用いることによって、必要最小限のメモリを用いるだけで、有音部分の削除によって音声が途切れてしまうことが少なくなり、内容が理解しやすくなる。第2の発明によれば、さらに、連続音声部分の密着によって異なる音に聞こえてしまうことが少なくなる。

【図面の簡単な説明】

【図1】

この発明の再生装置の一例を示す図である。

【図2】

音声処理部の一例を示す図である。

【図3】

振幅抑圧処理部の一例を示す図である。

【図4】

図3の例の振幅抑圧処理の説明に供する図である。

【図5】

図3の例の制御部が行う処理ルーチンの一例を示す図である。

【図6】

図3の例の振幅抑圧処理前と処理後の音声信号波形の一例を示す図である。

【図 7】

振幅抑圧処理部の他の例を示す図である。

【図 8】

振幅抑圧処理部のさらに他の例を示す図である。

【図 9】

図 7 および図 8 の例のバンドパスフィルタの説明に供する図である。

【図 1 0】

図 7 および図 8 の例の制御部が行う処理ルーチンの一例を示す図である。

【図 1 1】

図 7 および図 8 の例の振幅抑圧処理の説明に供する図である。

【図 1 2】

図 7 および図 8 の例の振幅抑圧処理前と処理後の音声信号波形の一例を示す図である。

【図 1 3】

速度変換処理の一例を示す図である。

【図 1 4】

有音部分と無音部分の識別判定の一例を示す図である。

【図 1 5】

出力バッファの入出力ポインタの動きを示す図である。

【図 1 6】

速度変換処理ルーチンの一例を示す図である。

【図 1 7】

データ圧縮の一例を示す図である。

【図 1 8】

データ圧縮の際に位置調整を行わない場合を示す図である。

【図 1 9】

データ圧縮の際に位置調整を行う場合の調整方法を示す図である。

【図 2 0】

無音部分削除の際のノイズの発生を示す図である。

【図 2 1】

無音部分削除の際のノイズ発生防止方法を示す図である。

【図 2 2】

音声処理部の他の例を示す図である。

【図 2 3】

音程対応処理単位期間設定処理ルーチンの一例を示す図である。

【図 2 4】

音声処理部のさらに他の例を示す図である。

【図 2 5】

音声レベル対応処理単位期間設定処理ルーチンの一例を示す図である。

【図 2 6】

音声レベルが低い点を速度変換処理の処理単位期間の区切れ目とする場合の例を示す図である。

【図 2 7】

音声処理部のさらに他の例を示す図である。

【図 2 8】

多チャンネル再生の場合の再生装置の一例を示す図である。

【図 2 9】

定常再生対応の再生装置の一例を示す図である。

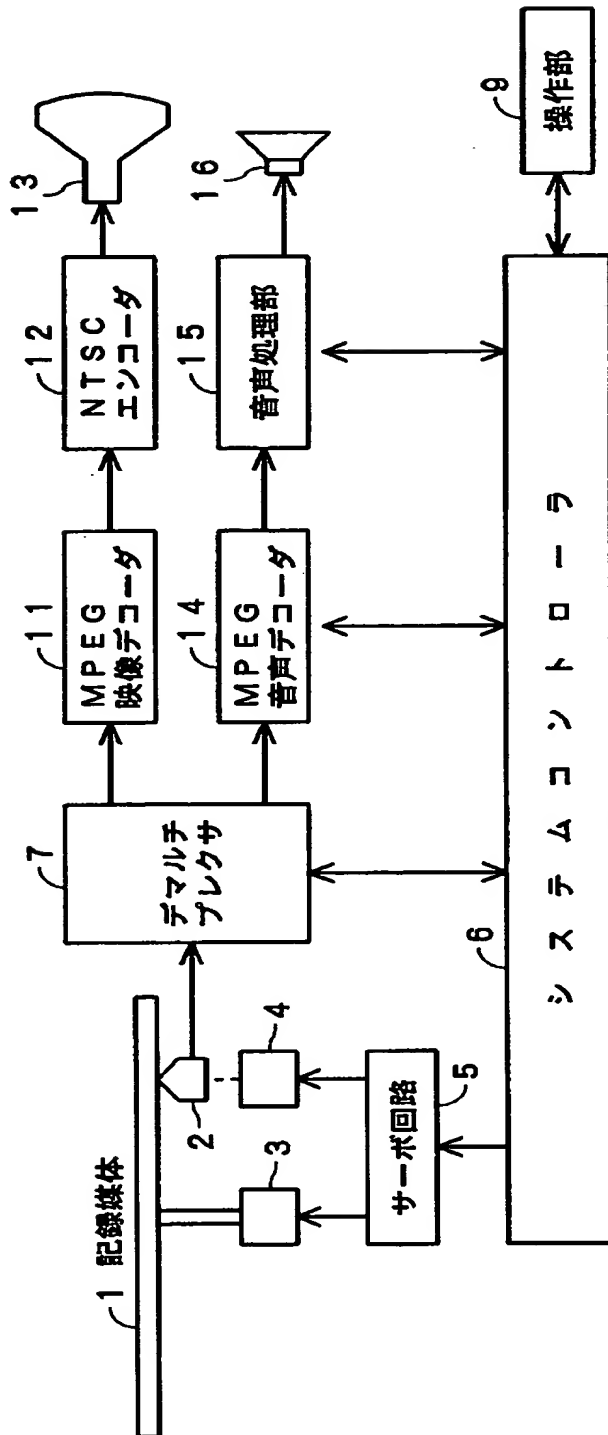
【符号の説明】

主要部については図中に全て記述したので、ここでは省略する。

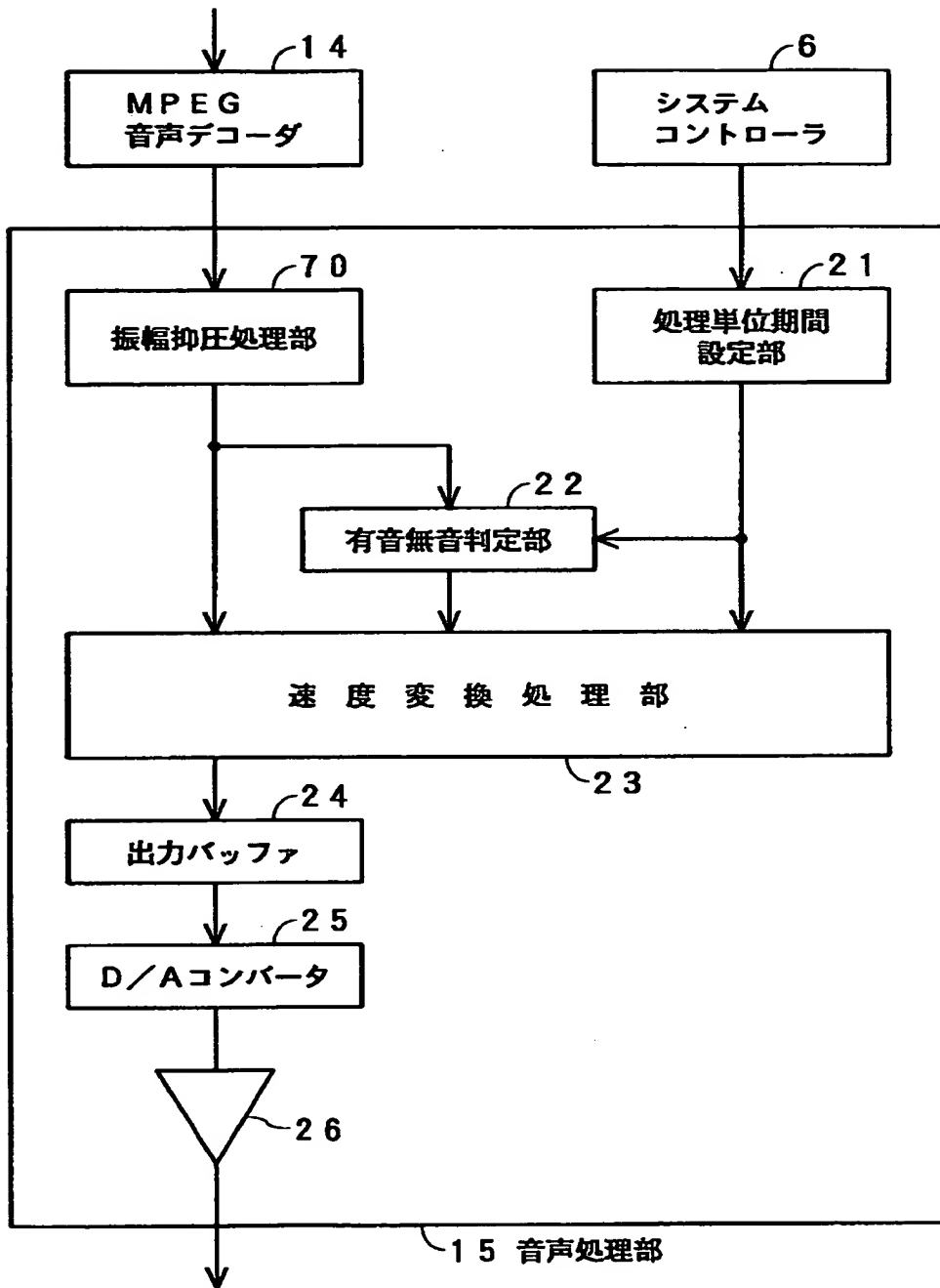


【書類名】 図面

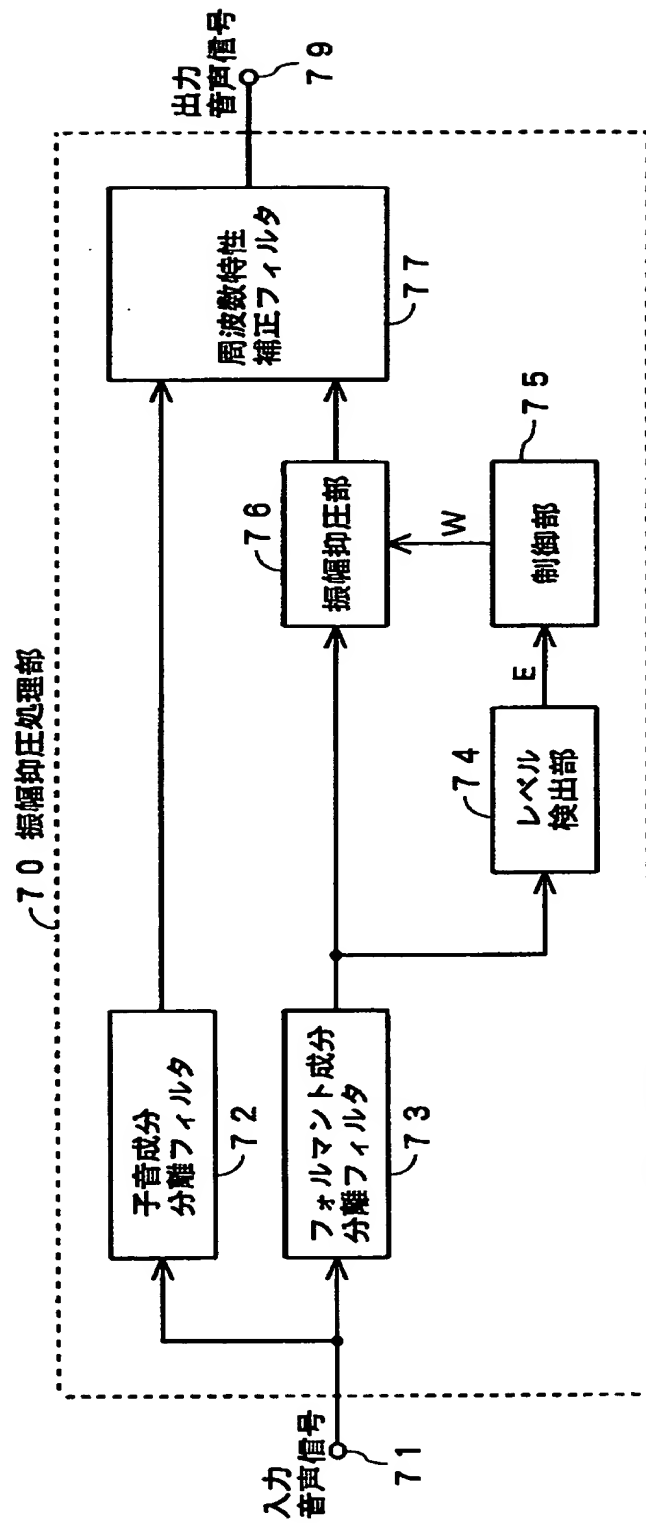
【図1】



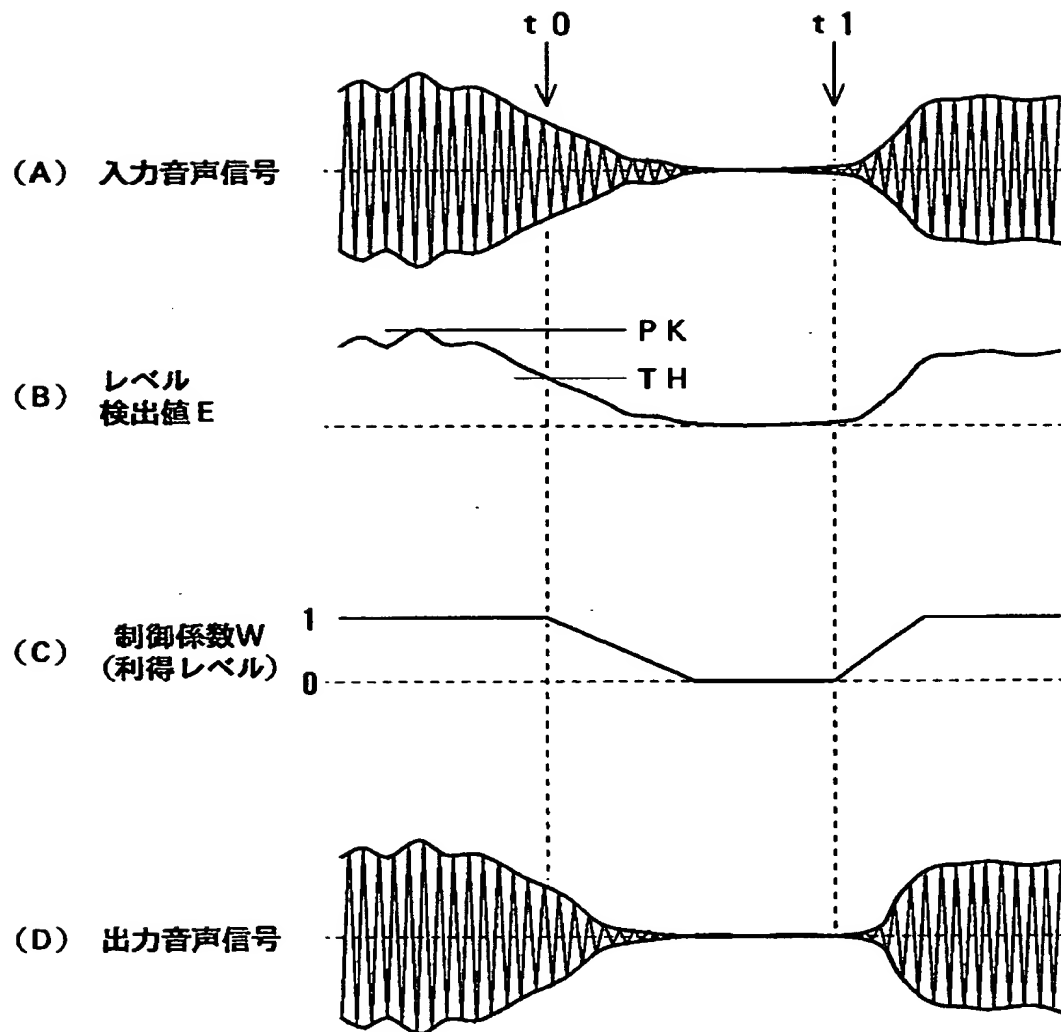
【図 2】



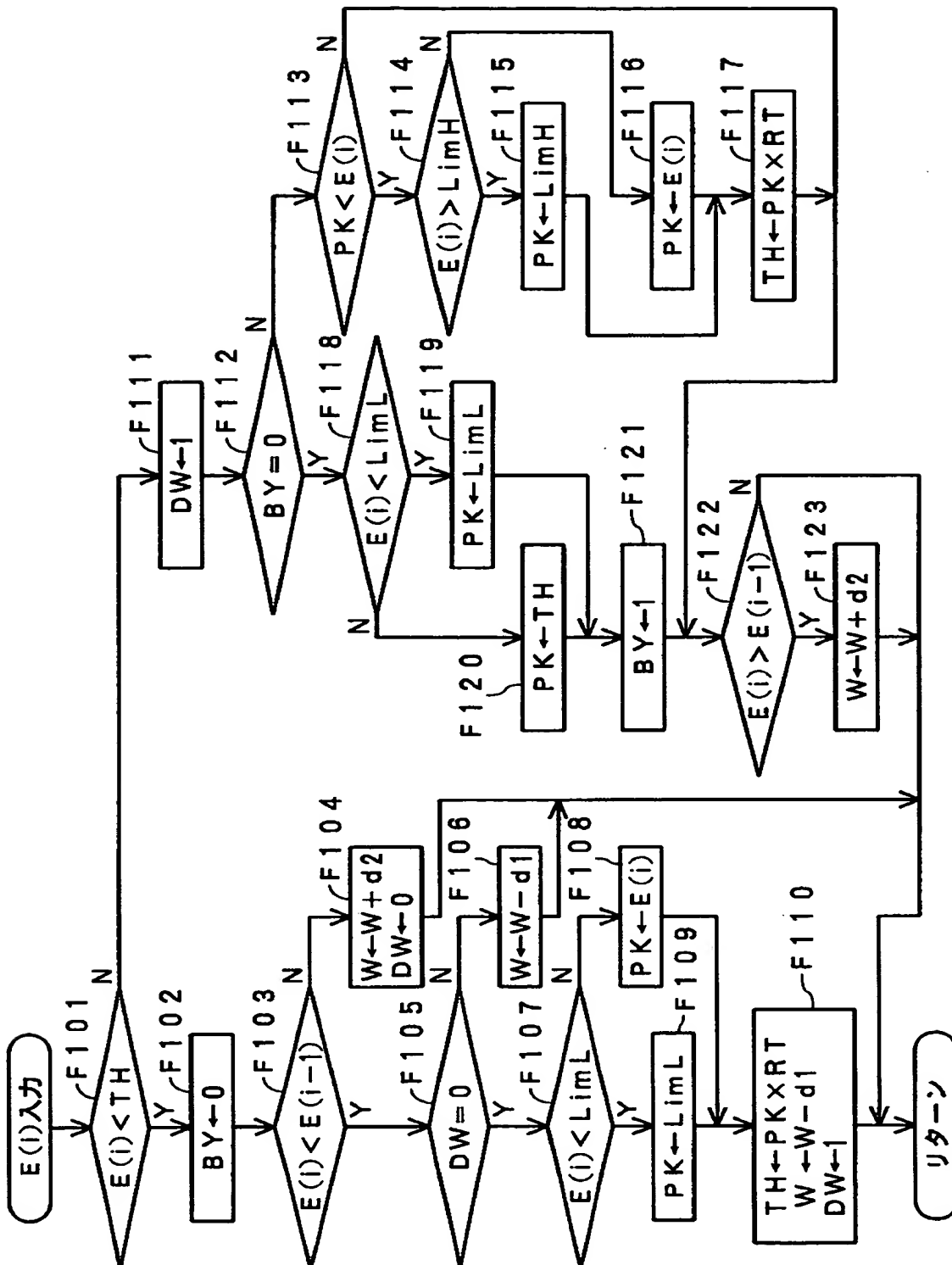
【図 3】



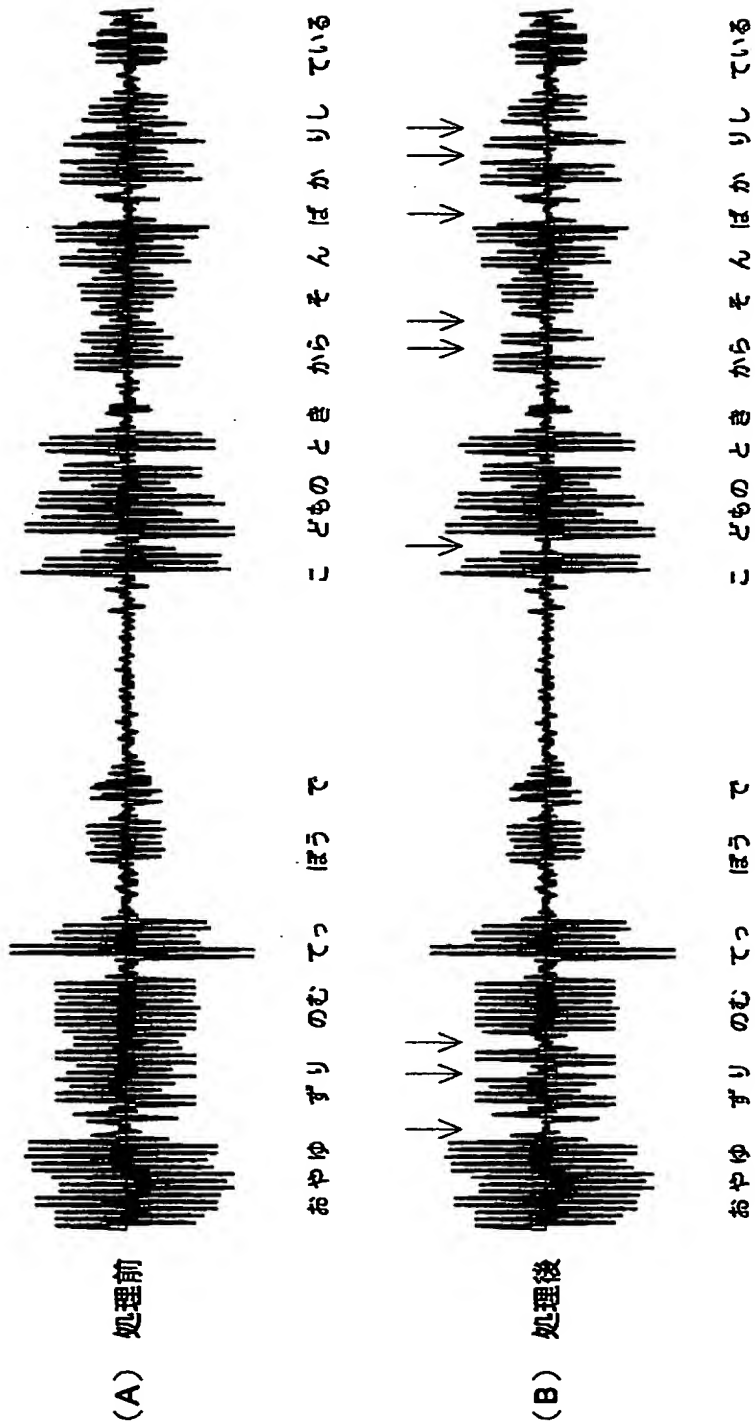
【図 4】



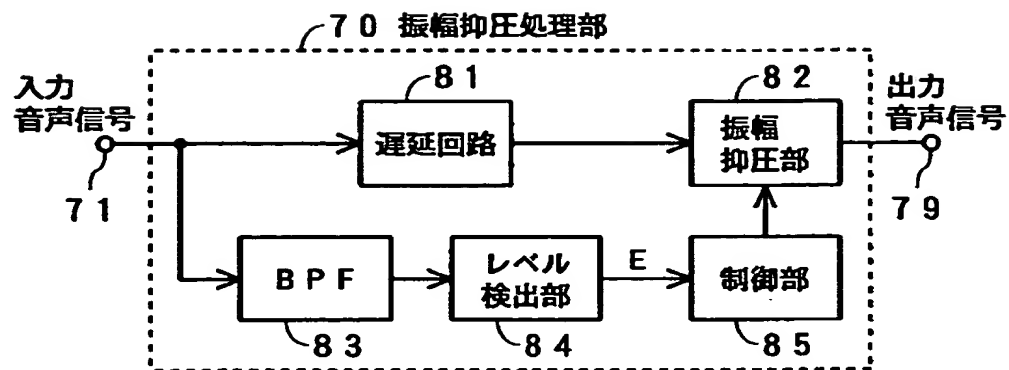
【図 5】



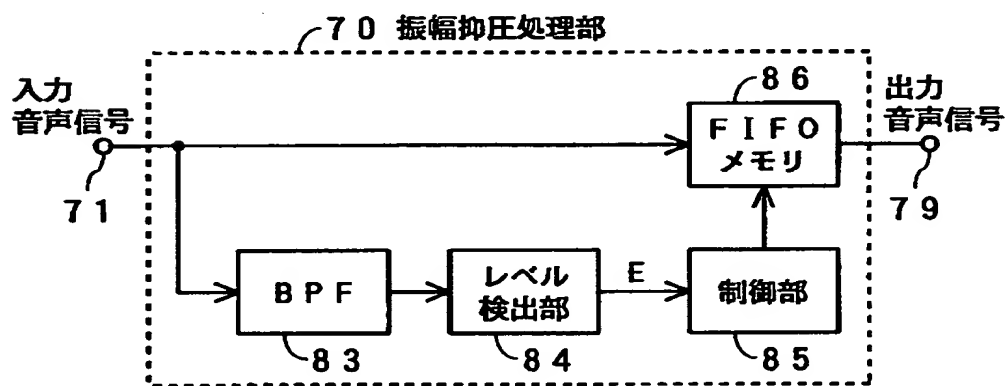
【図6】



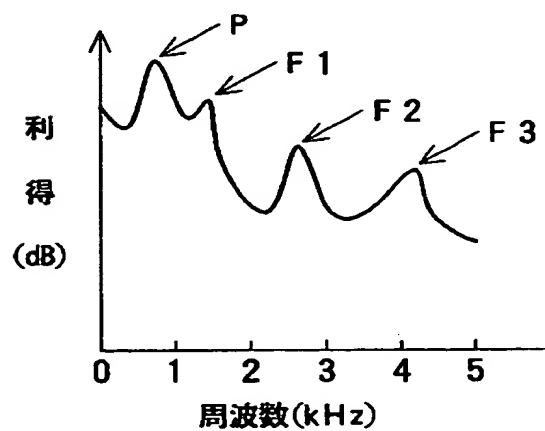
【図 7】



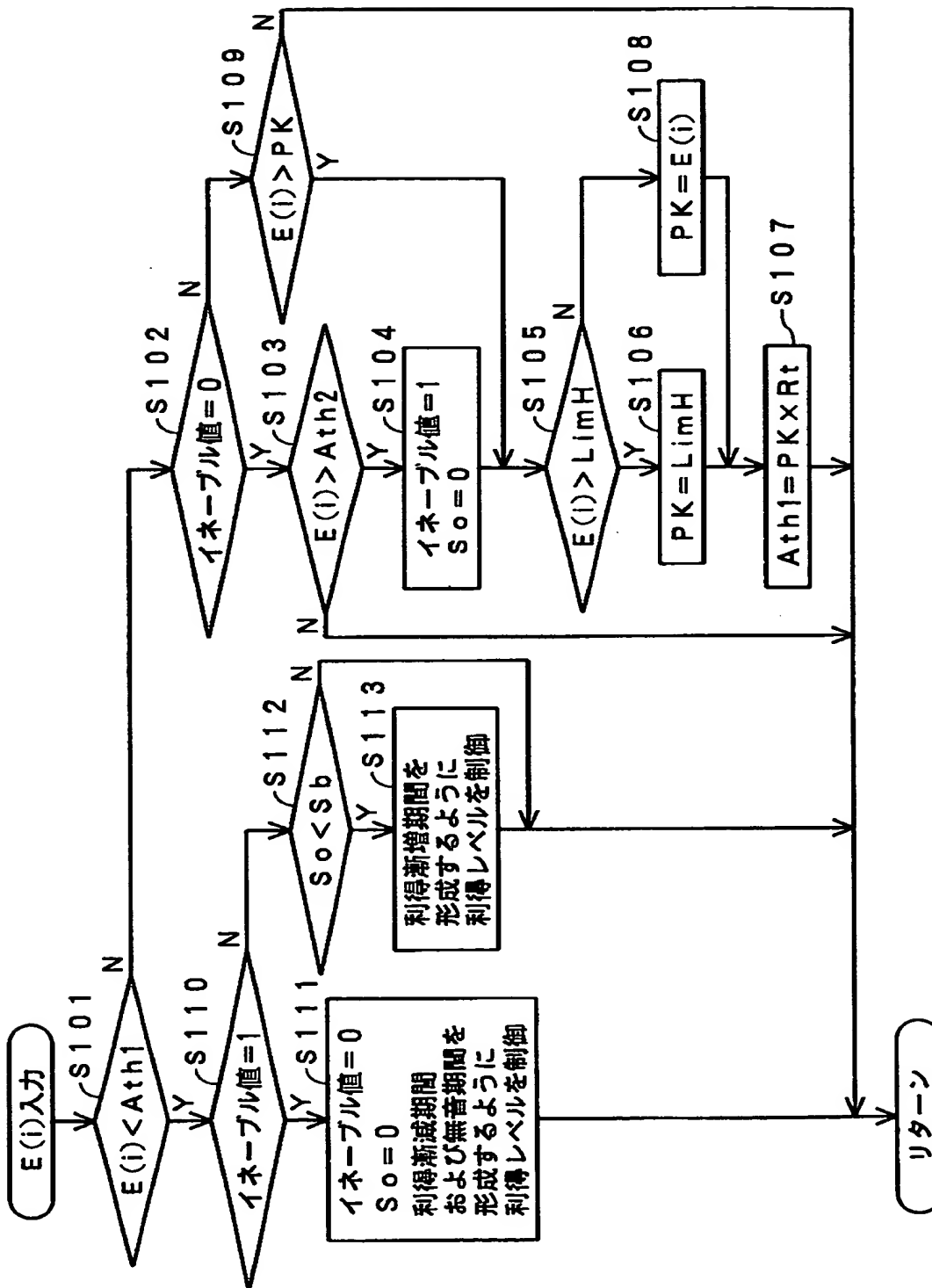
【図 8】



【図 9】

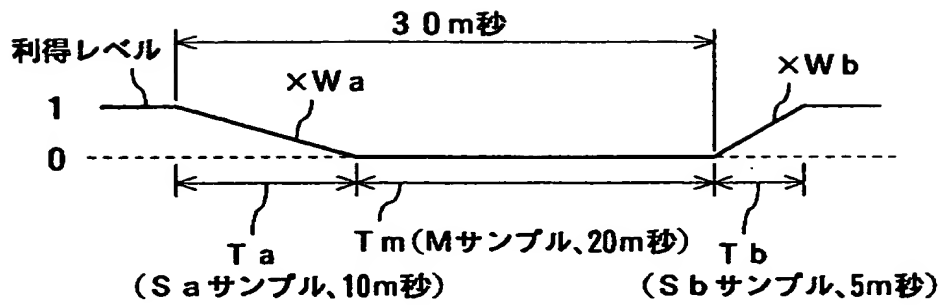


【図10】



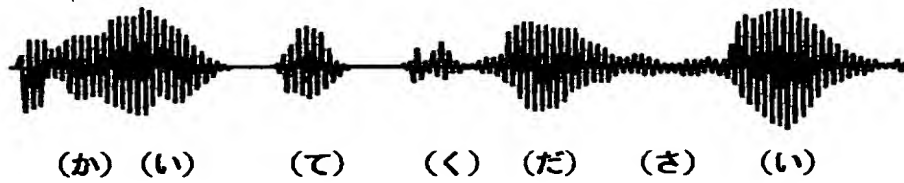


【図11】

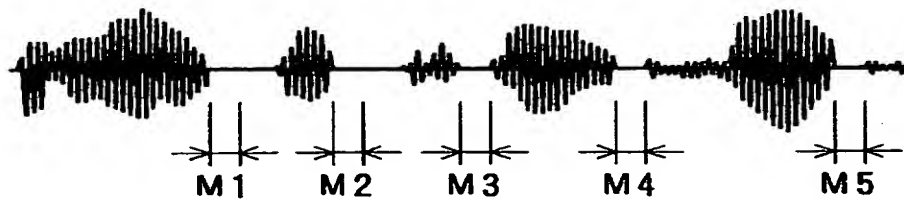


【図12】

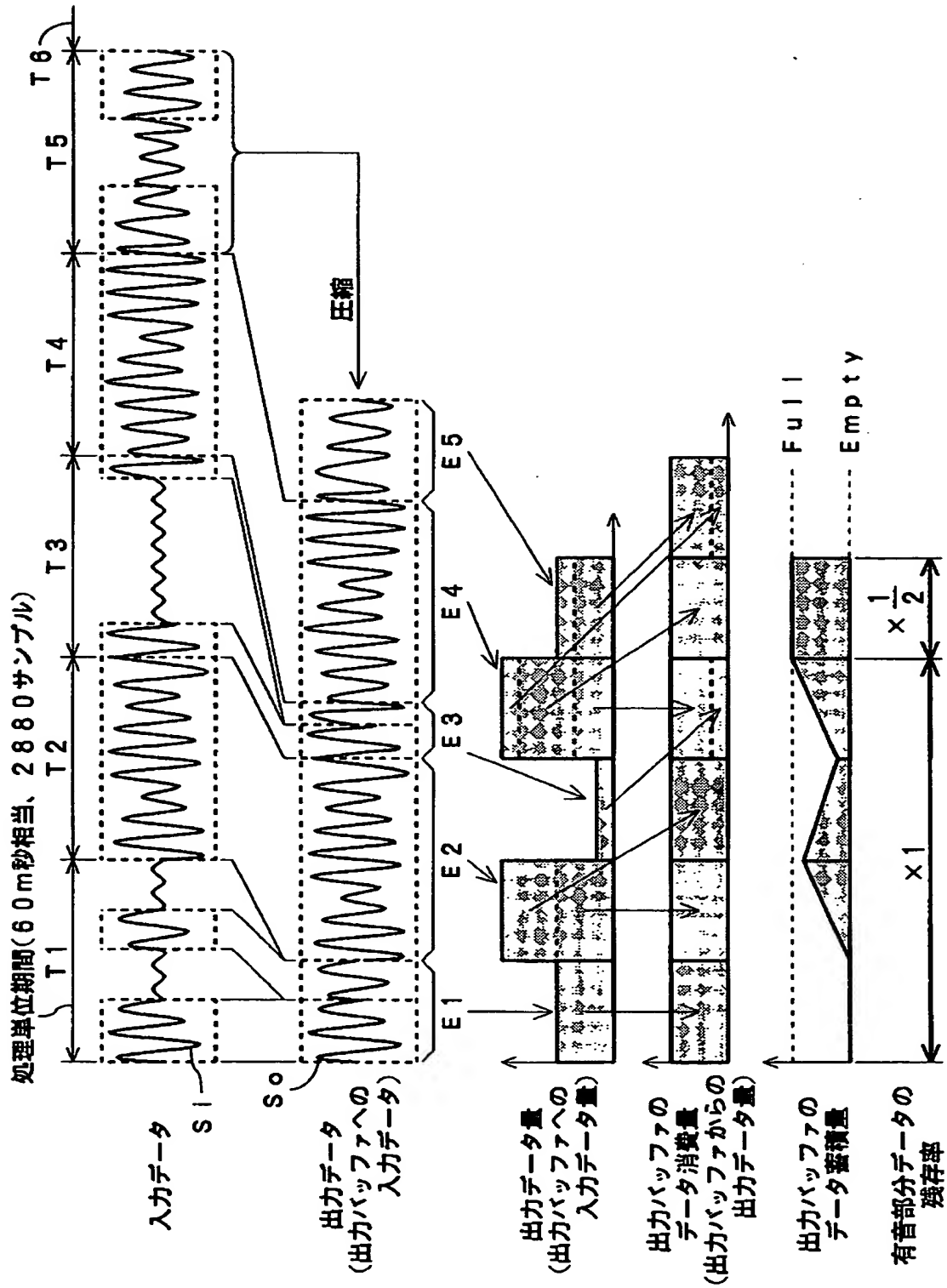
(A) 入力音声信号



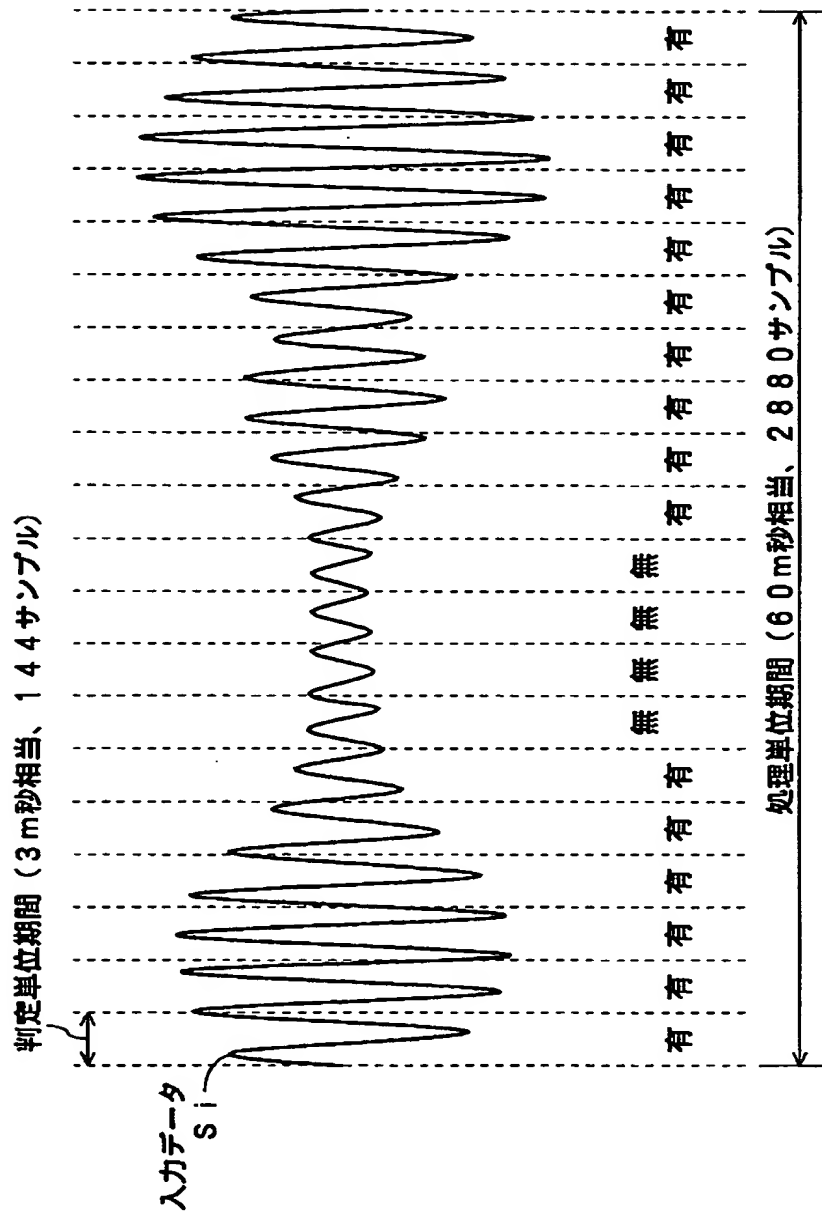
(B) 出力音声信号



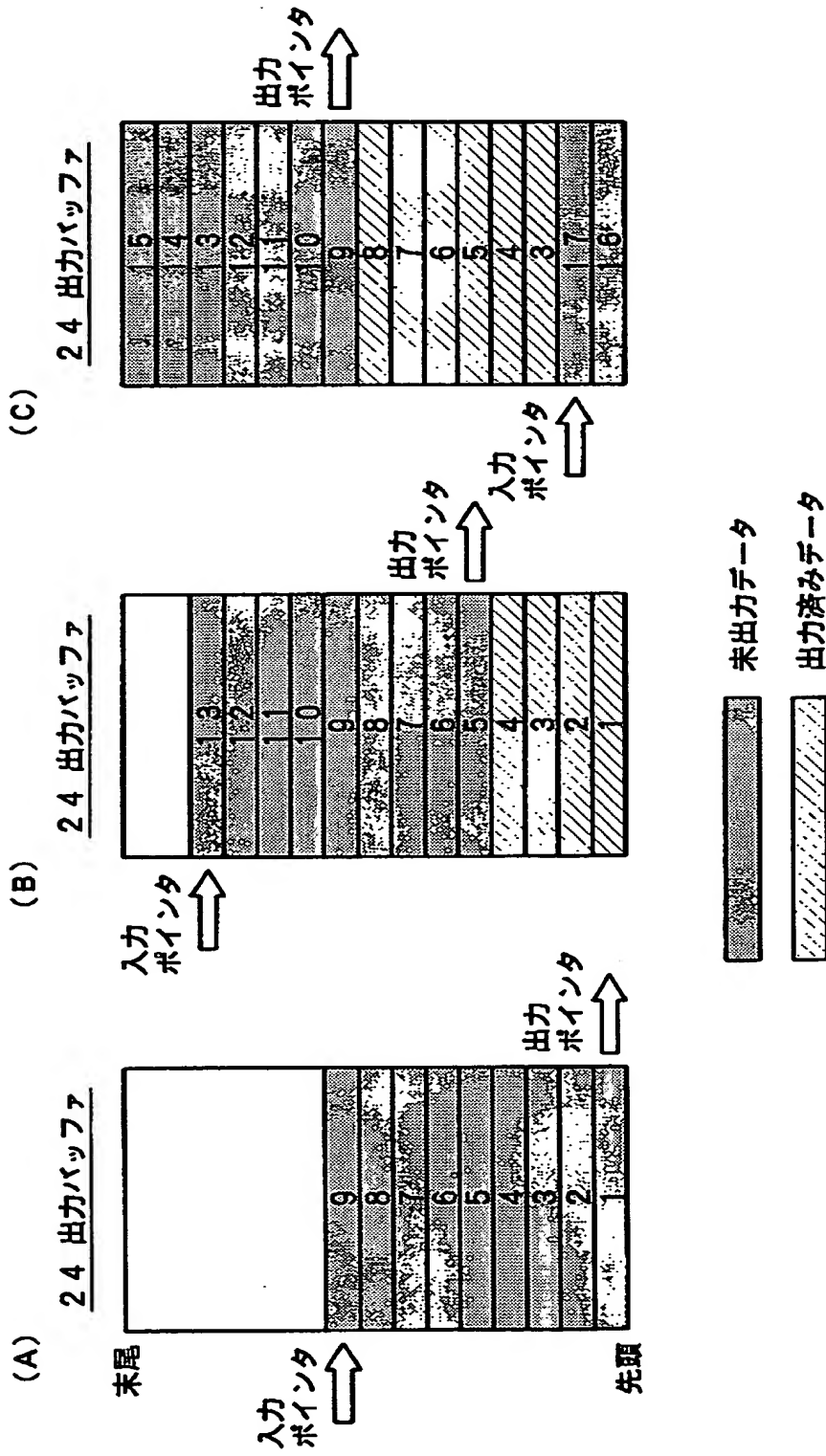
【図 13】



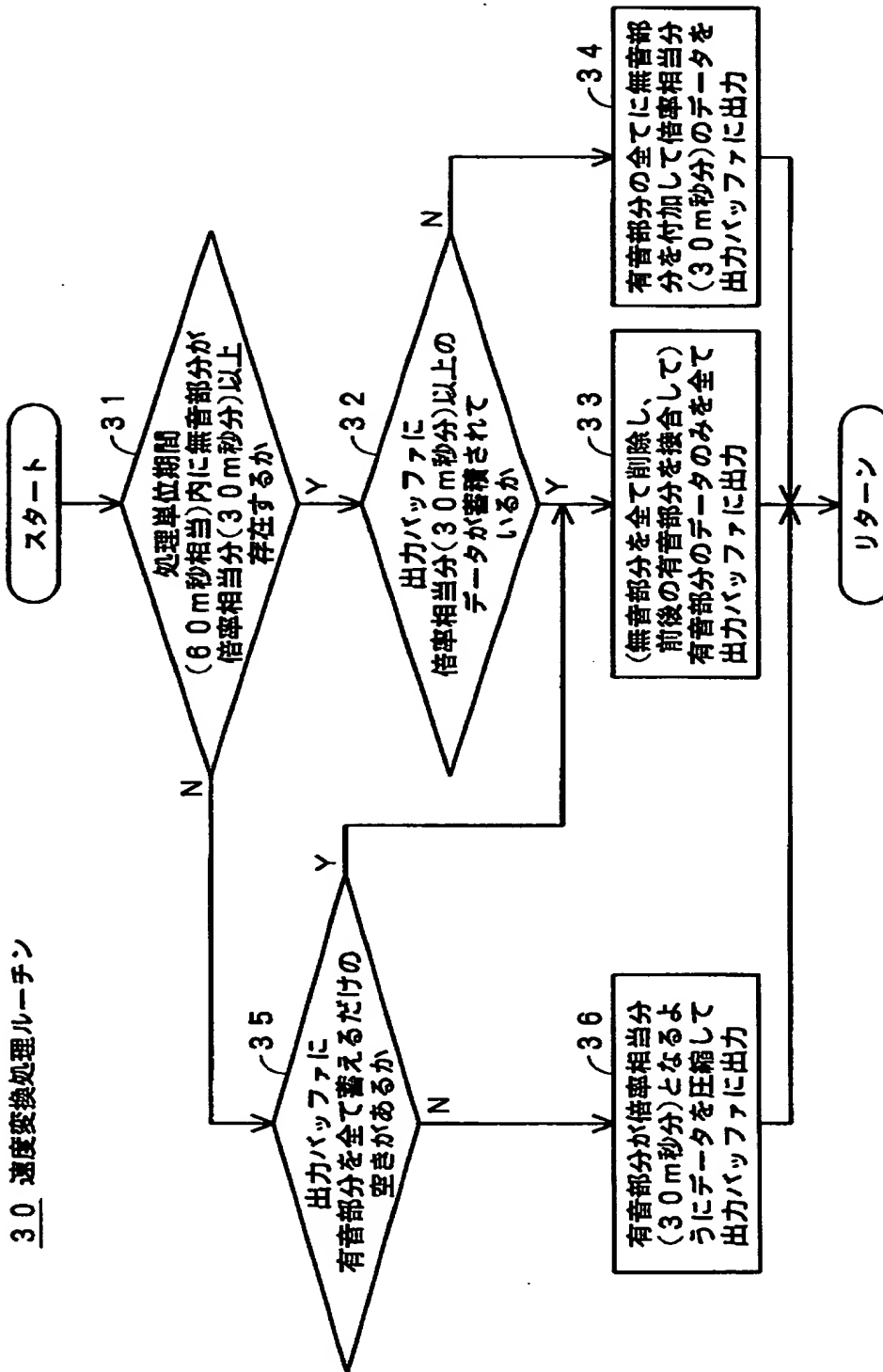
【圖 14】



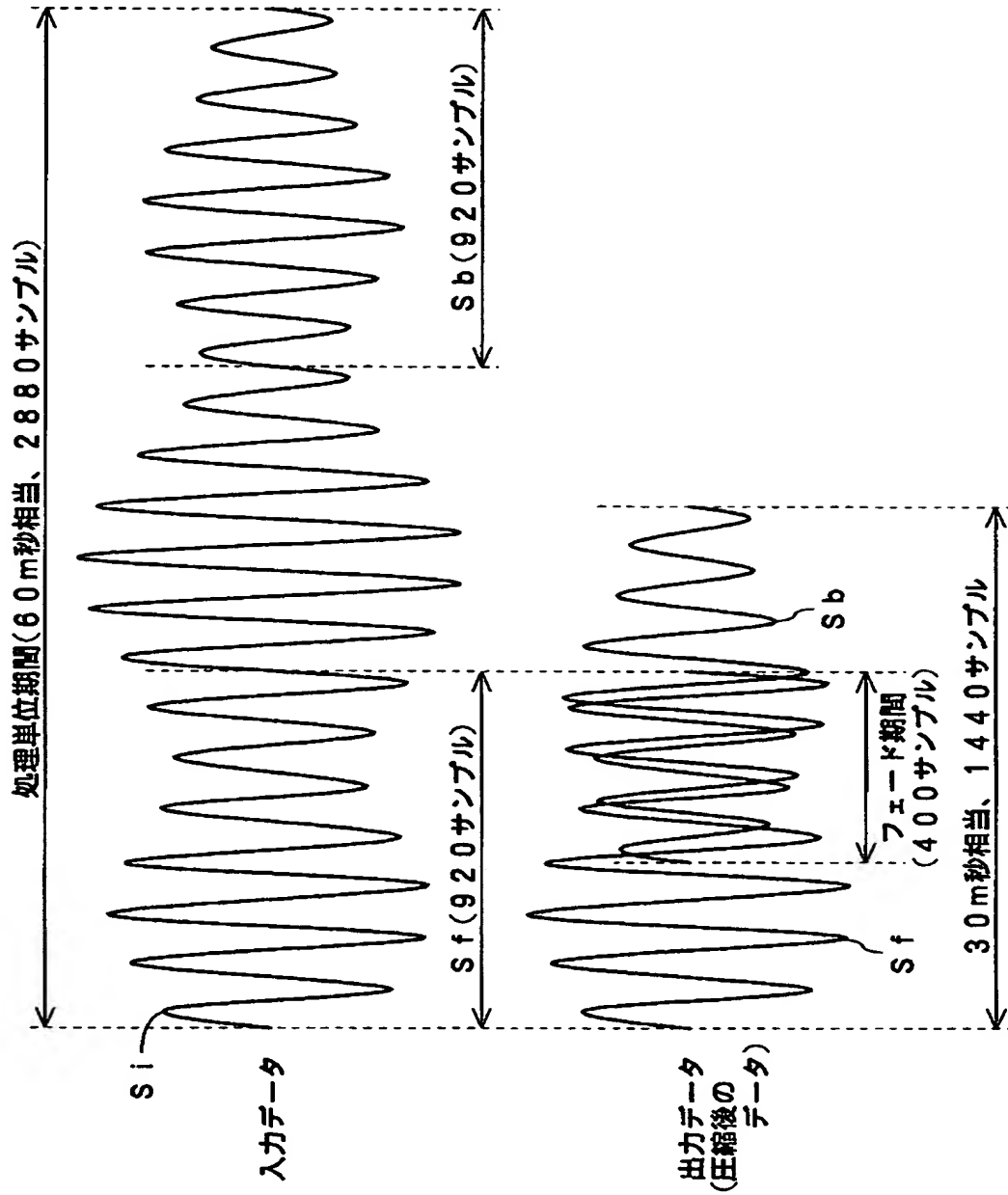
【図 15】



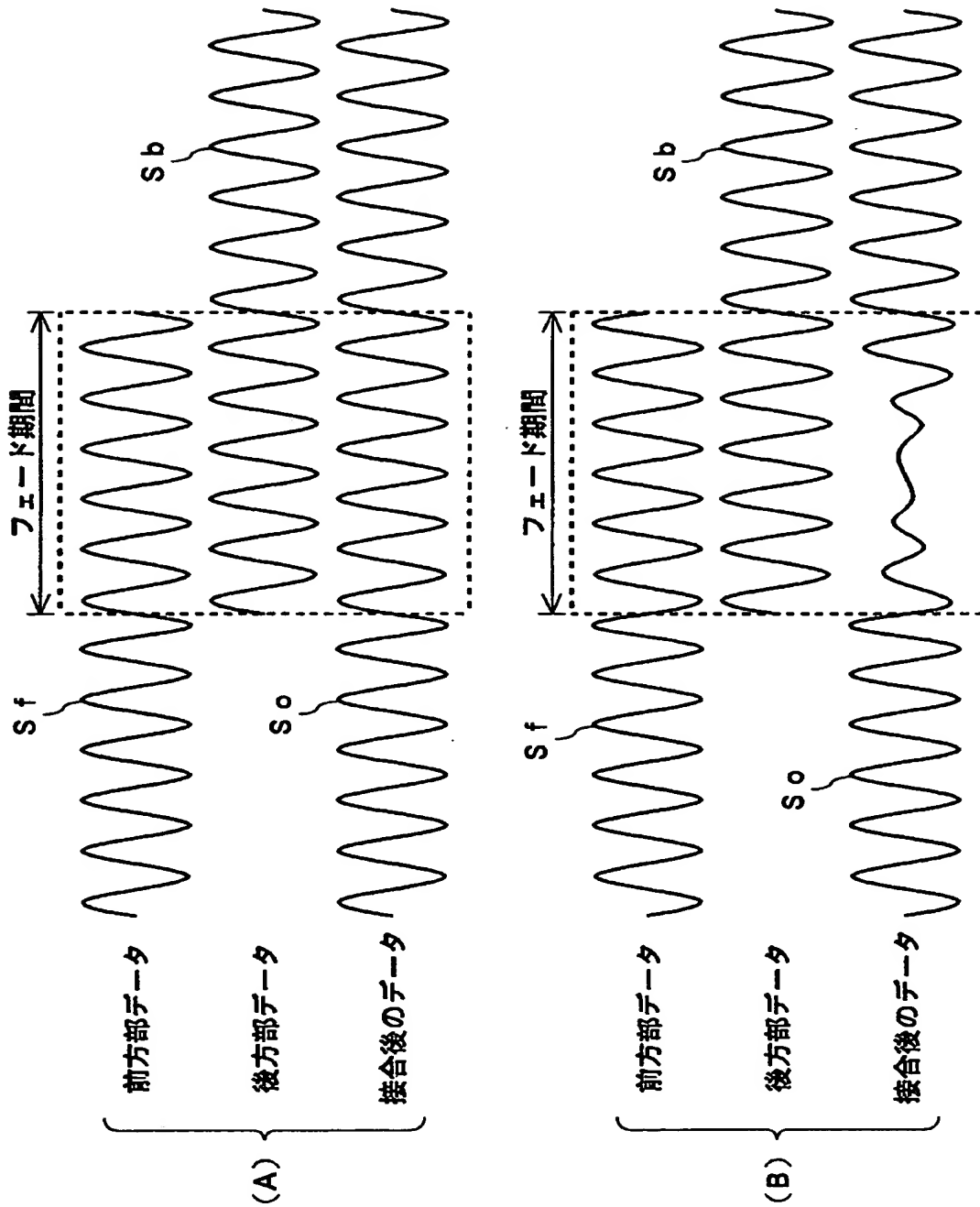
【図 16】



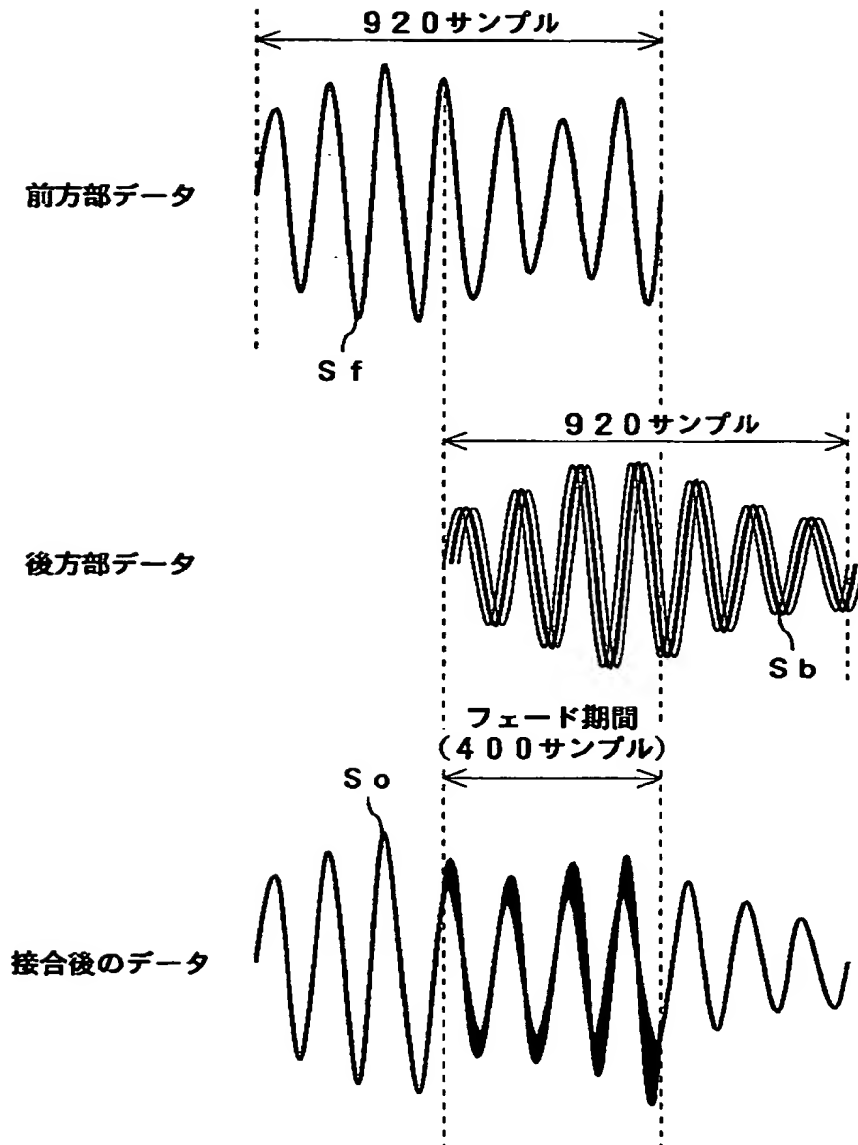
【図 17】



【図 18】

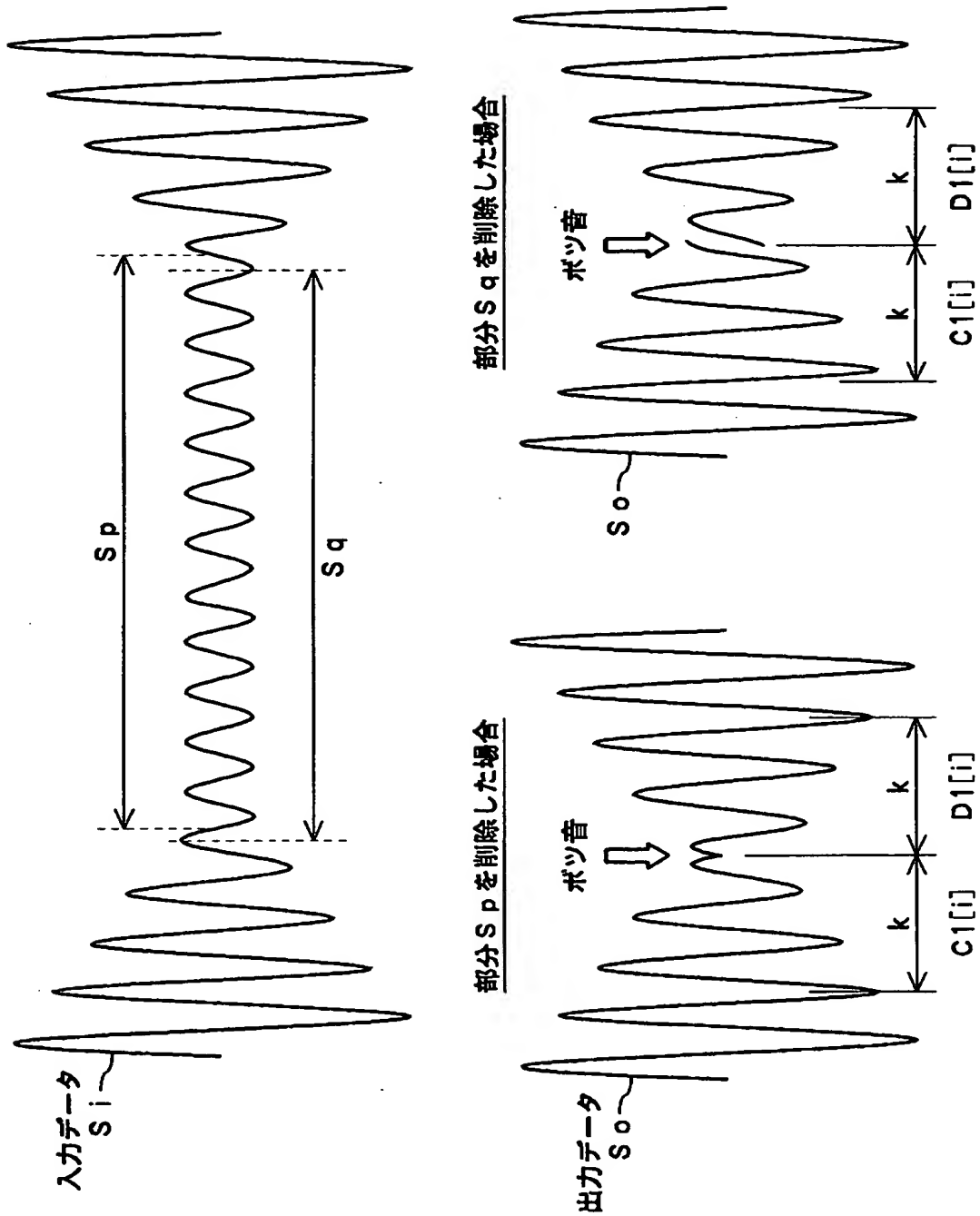


【図 1 9】

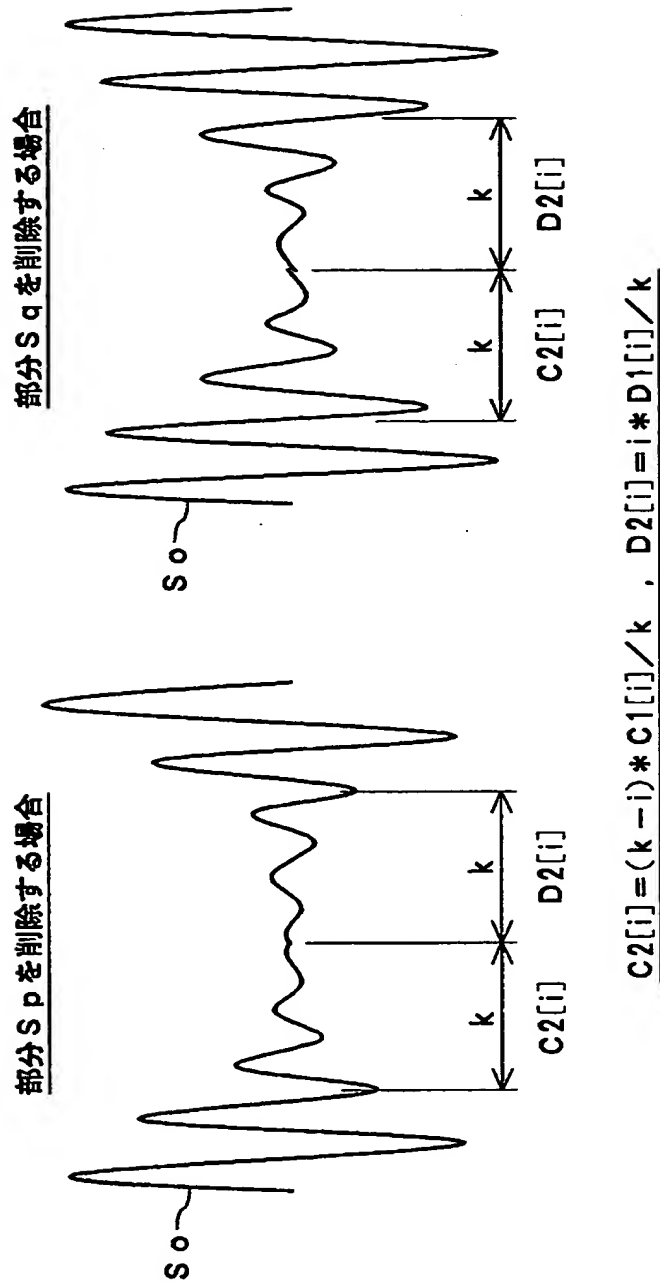




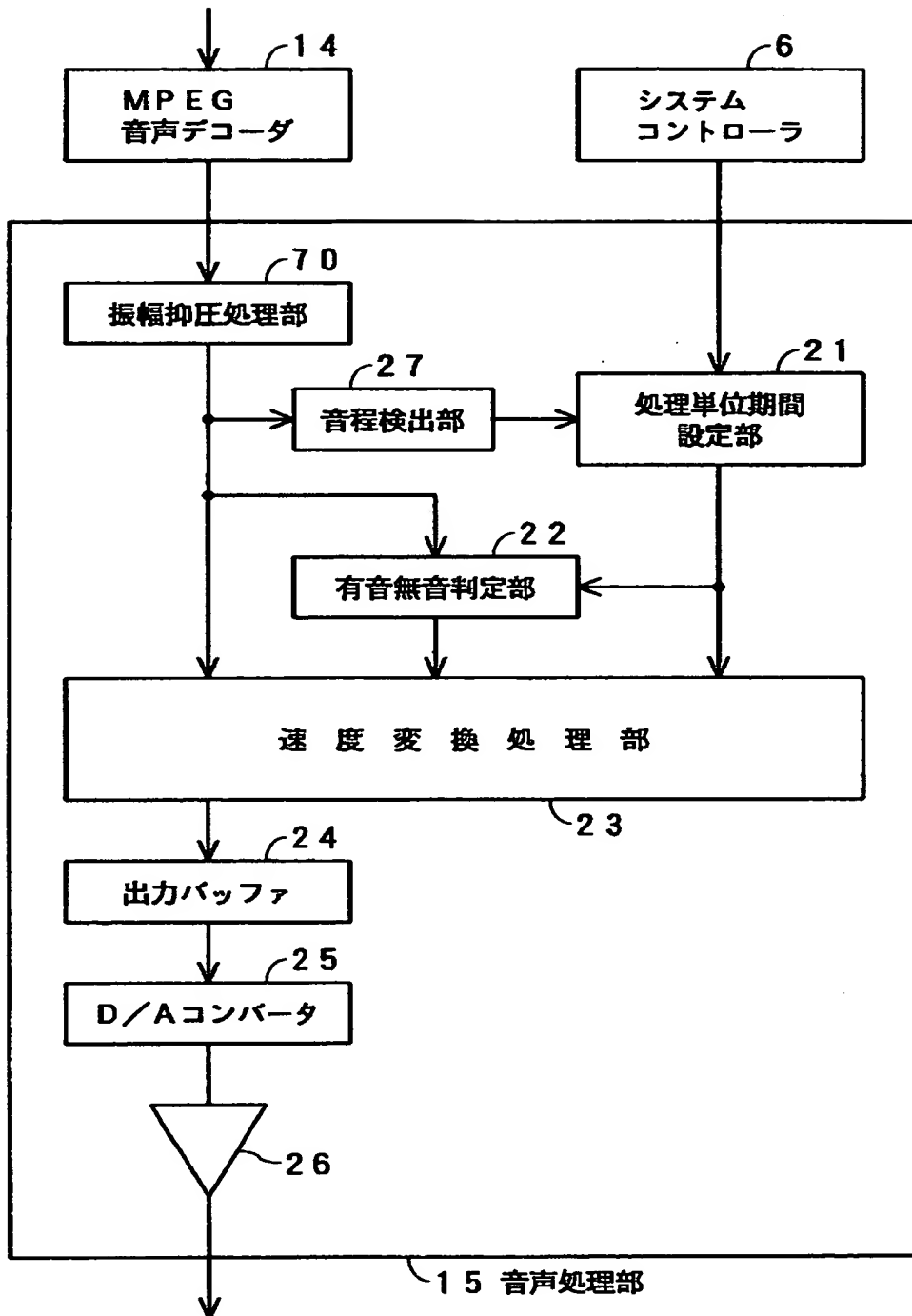
【図 20】



【図 2 1】

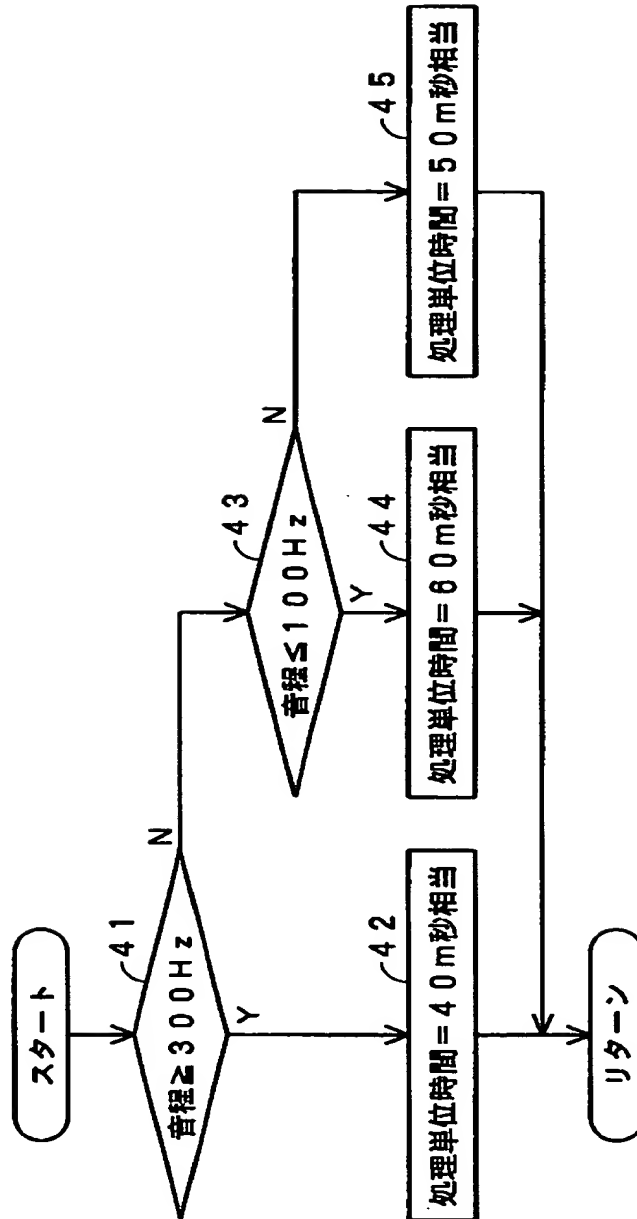


【図 22】

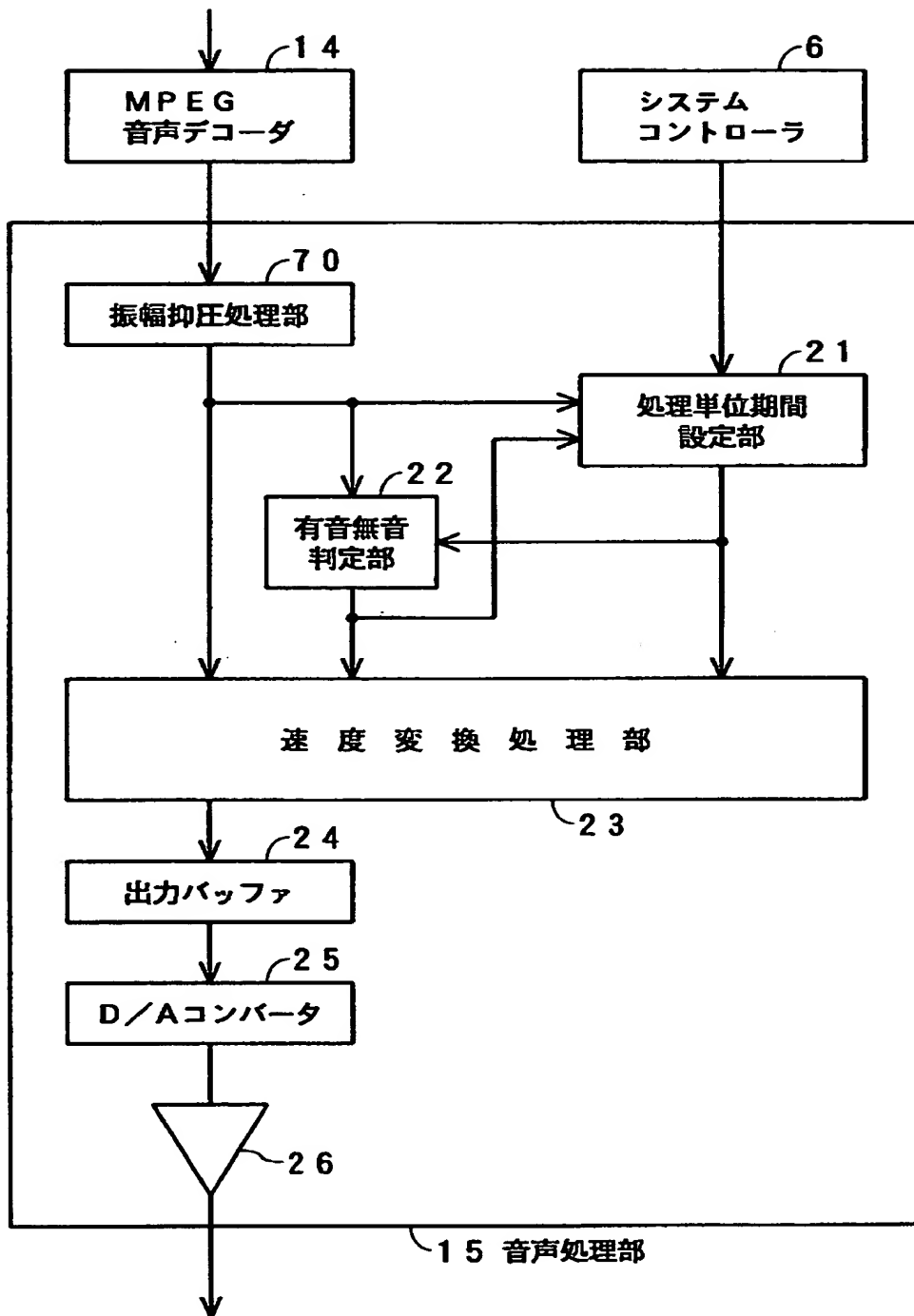


【図 2 3】

4 0 音程対応処理単位期間設定処理ルーチン

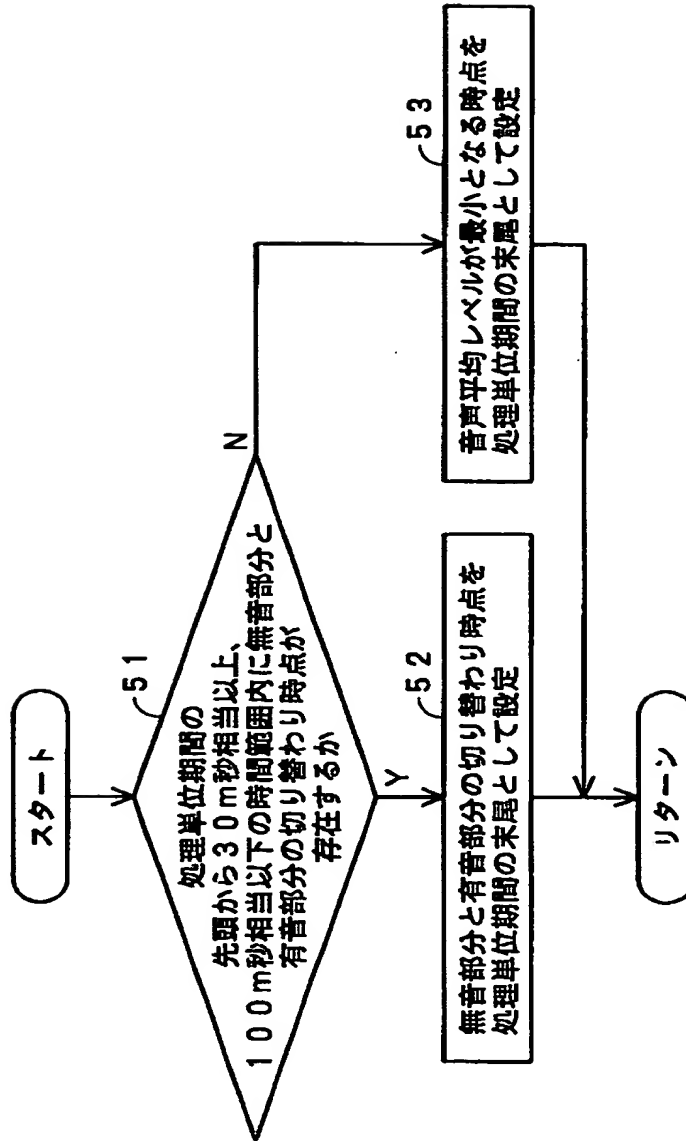


【図 2 4】

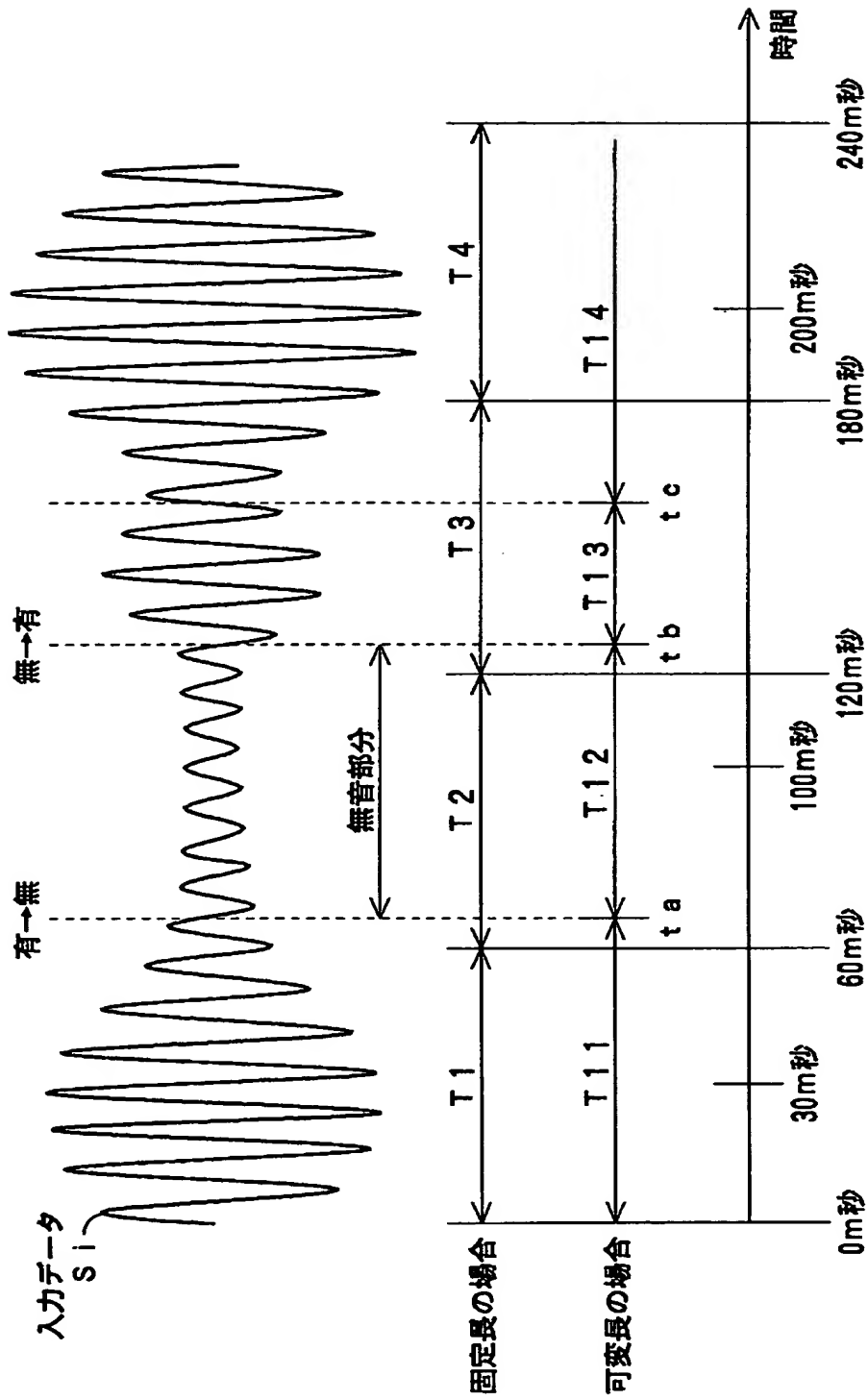


【図 2 5】

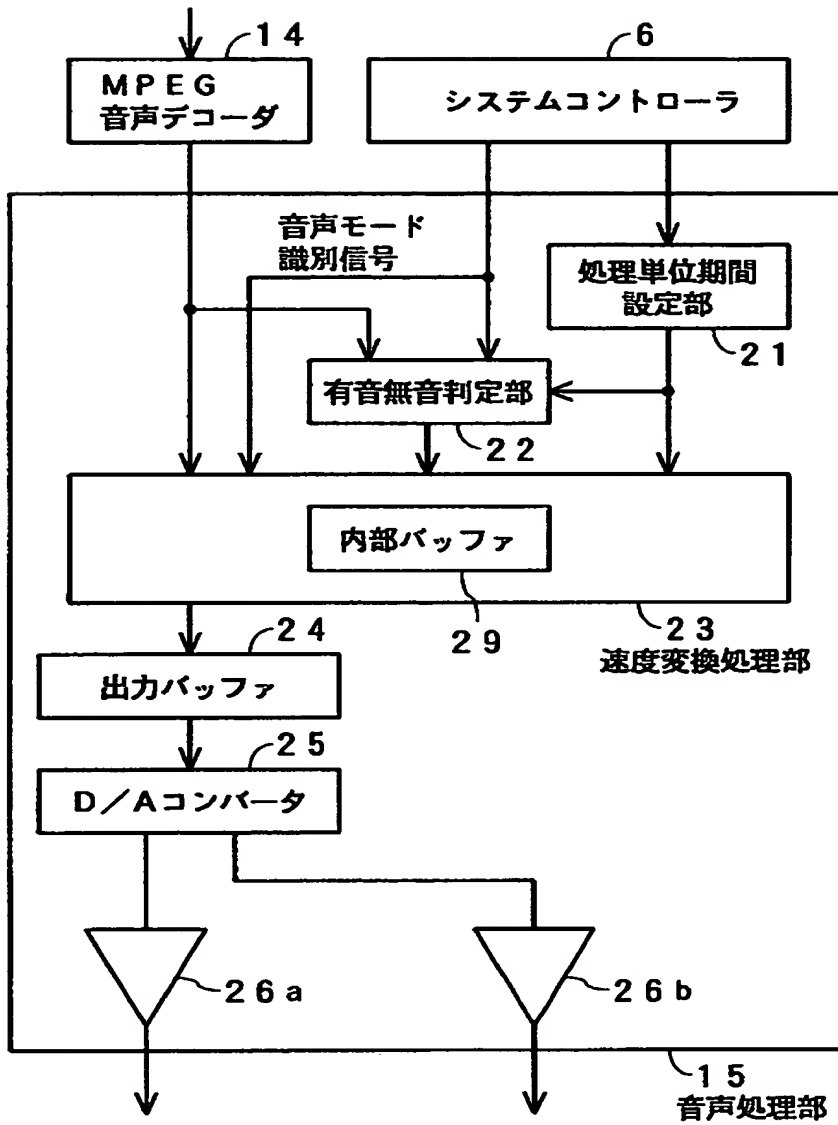
5 0 音声レベル対応処理単位期間設定処理ルーチン



【図 2 6】

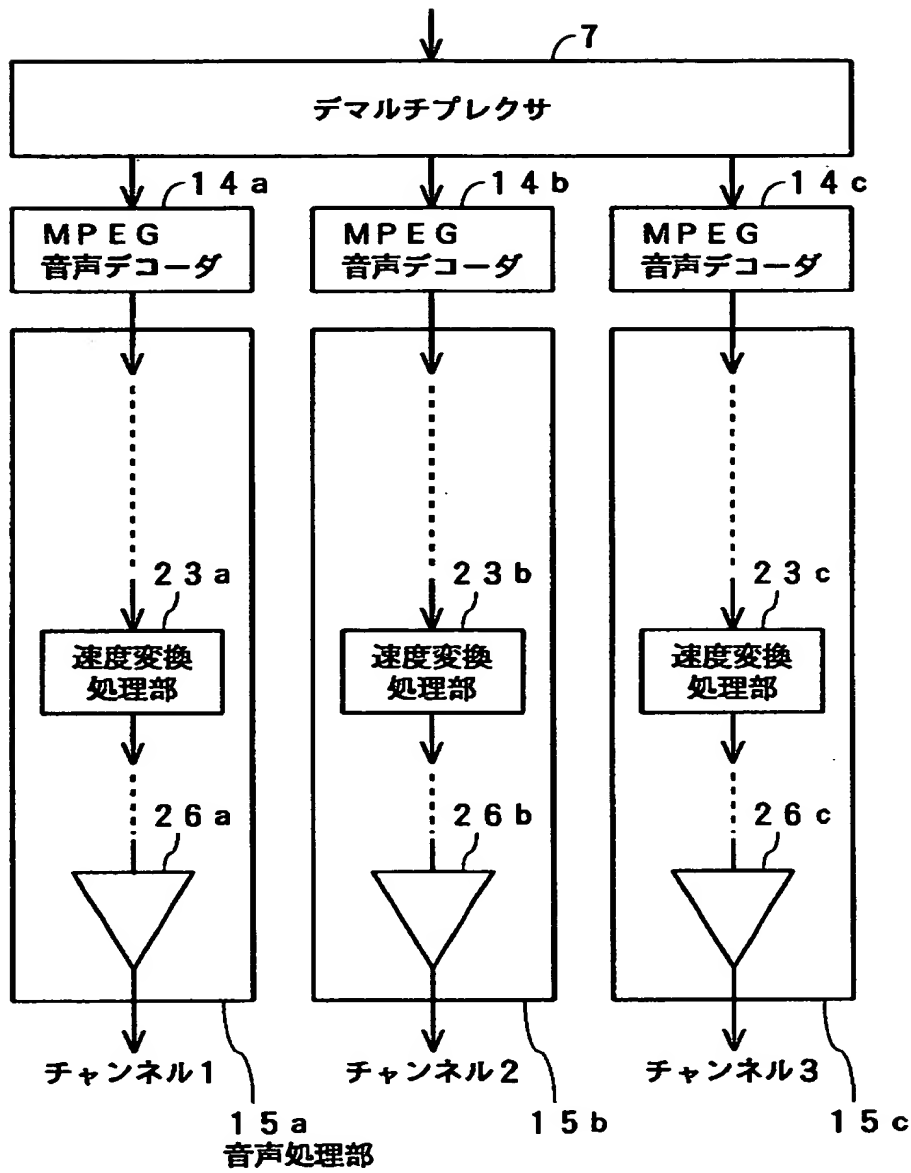


【図 2 7】

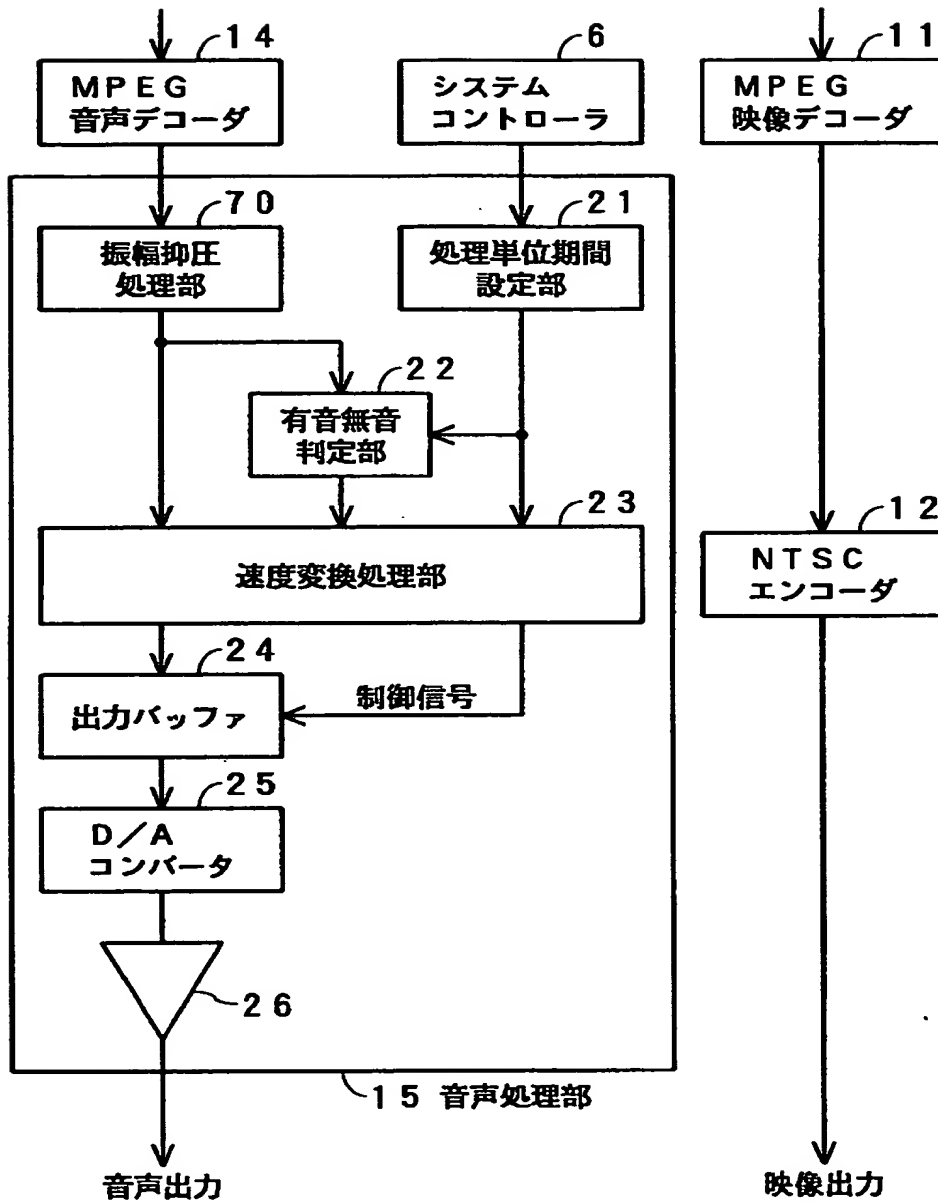




【図 2 8】



【図 2 9】



【書類名】 要約書

【要約】

【課題】 記録媒体から音声信号を定常速度より速い速度で再生し、通常の音程で出力する方法で、音声途切れることが少なく、内容が理解しやすくなるようにする。

【解決手段】 2倍速再生によって得られた元の再生音声信号である入力データ  $S_i$  を 60 m 秒相当の処理単位期間  $T_1, T_2 \dots$  ごとに区切る。期間  $T_1$  では、有音部分と無音部分が 30 m 秒分ずつ存在し、出力バッファにデータが全く蓄積されていないので、無音部分を全て削除し、前後の有音部分を接合して、有音部分のデータのみを全て出力バッファに書き込む。期間  $T_2$  では、60 m 秒分のデータが全て有音部分であり、出力バッファに 60 m 秒分のデータを全て蓄えるだけの空きがあるので、有音部分のデータをそのまま出力バッファに書き込む。期間  $T_2$  で有音部分が 30 m 秒分未満しか存在しないときには、有音部分の全てに無音部分の一部を付加して、30 m 秒分のデータを出力バッファに書き込む。期間  $T_5$  では、60 m 秒分のデータが全て有音部分であり、出力バッファに 60 m 秒分のデータを全て蓄えるだけの空きがないので、60 m 秒分の有音部分のデータを 30 m 秒分に圧縮して出力バッファに書き込む。

【選択図】 図 1 3

認定・付加情報

特許出願の番号	特願2000-171556
受付番号	50000710548
書類名	特許願
担当官	第八担当上席 0097
作成日	平成12年 6月13日

<認定情報・付加情報>

【特許出願人】

【識別番号】	000002185
【住所又は居所】	東京都品川区北品川6丁目7番35号
【氏名又は名称】	ソニー株式会社

【代理人】

申請人	
【識別番号】	100091546
【住所又は居所】	東京都新宿区西新宿8丁目12番1号 篠ビル8 階 佐藤正美特許事務所
【氏名又は名称】	佐藤 正美

出 願 人 履 歴 情 報

識別番号                   【000002185】

1. 変更年月日	1990年 8月30日
[変更理由]	新規登録
住 所	東京都品川区北品川6丁目7番35号
氏 名	ソニー株式会社